CANADIAN JOURNAL OF MATHEMATICS

Journal Canadien de Mathématiques

VOLUME I

1949

Published for

THE CANADIAN MATHEMATICAL CONGRESS

by the University of Toronto Press

Math Lib
direct

INDEX OF VOLUME I

JO

JOURNAL OF MATHEMATICS

Journal Canadien de Mathématiques

VOL. I - NO. I

1949

Published for

THE CANADIAN MATHEMATICAL CONGRESS

by the University of Toronto Press

The chief languages of the *Journal* are English and French.

Manuscripts for publication in the *Journal* should be sent to the *Editor-in-Chief*, H. S. M. Coxeter, University of Toronto. Every paper should contain an introduction summarizing the results as far as possible in such a way as to be understood by the non-expert.

All other correspondence should be addressed to the *Managing Editor*, G. de B. Robinson, University of Toronto.

The *Journal* is published quarterly. Subscriptions should be sent to the *Managing Editor*. The price per volume of four numbers is $6.00. This is reduced to $3.00 for individuals who are members of the following Societies:

Canadian Mathematical Congress
American Mathematical Society
Mathematical Association of America
London Mathematical Society
Société Mathématique de France

The Canadian Mathematical Congress gratefully acknowledges the assistance of the following towards the cost of publishing this Journal:

University of British Columbia
University of Manitoba
McGill University
Queen's University
University of Toronto
and
The American Mathematical Society

RF

ON THE PRODUCT OF TWO POWER SERIES

H. DAVENPORT AND G. PÓLYA

WE consider the product of two power series with positive coefficients:

$$(\Sigma u_n x^n)(\Sigma v_n x^n) = \Sigma w_n x^n.$$

What conditions will ensure that the coefficients w_n shall be either (i) monotonic, or (ii) logarithmically convex? By the latter, we mean that $w_n^2 \leq w_{n-1}w_{n+1}$ for $n = 1, 2, \ldots$. In investigating this question, which was suggested by a special example, we have found it convenient to express the conditions in terms of the ratios of u_n, v_n to certain binomial coefficients, rather than in terms of u_n, v_n themselves.

We introduce α and β such that

$$\alpha > 0, \quad \beta > 0, \quad \alpha + \beta = 1 \tag{1}$$

and let

$$\alpha_n = \frac{\alpha(\alpha+1)\ldots(\alpha+n-1)}{1\,.\,2\ldots n}, \quad \beta_n = \frac{\beta(\beta+1)\ldots(\beta+n-1)}{1\,.\,2\ldots n} \tag{2}$$

for $n \geq 1$; $\alpha_0 = \beta_0 = 1$. Let

$$a_n = u_n/\alpha_n, \quad b_n = v_n/\beta_n$$

so that a_n and b_n are positive, and

$$w_n = \alpha_0 a_0 \beta_n b_n + \alpha_1 a_1 \beta_{n-1} b_{n-1} + \ldots + \alpha_n a_n \beta_0 b_0. \tag{3}$$

We have been led to the following very elementary results, which appear, however, to be new.

THEOREM 1. *If a_n and b_n are both monotonic increasing, so is w_n, and if a_n and b_n are both monotonic decreasing, so is w_n.*

THEOREM 2. *If a_n and b_n are both logarithmically convex, so is w_n.*

We prove these theorems in **1** and **2**, and add some general remarks concerning them in **3**. In **4** we apply them to the special example from which our investigation started. In **5** we mention the integral analogues.

1. The proof of Theorem 1 may be decomposed into two steps, the first of which is concerned only with properties of the binomial coefficients.

Put

$$\begin{cases} p_0 = \alpha_0\beta_n, & p_1 = \alpha_1\beta_{n-1}, \ldots, & p_n = \alpha_n\beta_0 \\ q_0 = \alpha_0\beta_{n+1}, & q_1 = \alpha_1\beta_n, \ldots, & q_{n+1} = \alpha_{n+1}\beta_0. \end{cases} \tag{4}$$

Then we assert that

$$p_0 + p_1 + \ldots + p_n = q_0 + q_1 + \ldots + q_{n+1} = 1, \tag{5}$$

and

$$q_0 < p_0 < q_0 + q_1 < p_0 + p_1 < \ldots < q_0 + q_1 + \ldots + q_n < p_0 + p_1 + \ldots p_n. \tag{6}$$

Thus we assert that the successive partial sums of the two sequences $p_0, p_1, \ldots$ and $q_0, q_1, \ldots$ separate each other. If we imagine each sequence represented

Received January 29, 1948.

by a row of blocks, the two rows will have a form similar to that of two neighbouring rows of tiles in a wall, and we can express the property in question by saying that the two sequences are "tilewise ordered."

Of the two results (5) and (6), the former is immediate, since, by (2),

$$\sum_0^\infty \alpha_n x^n = (1-x)^{-\alpha}, \; \sum_0^\infty \beta_n x^n = (1-x)^{-\beta}$$

and so, by (1),

$$\sum_0^\infty (\alpha_0\beta_n + \ldots + \alpha_n\beta_0)x^n = (1-x)^{-1} = \sum_0^\infty x^n.$$

To prove (6), we observe that, by (1) and (2), the α_n and β_n are monotonic decreasing, whence

$$\begin{aligned} q_0 + q_1 + \ldots + q_k &= \alpha_0\beta_{n+1} + \alpha_1\beta_n + \ldots + \alpha_k\beta_{n+1-k} \\ &< \alpha_0\beta_n + \alpha_1\beta_{n-1} + \ldots + \alpha_k\beta_{n-k} \\ &= p_0 + p_1 + \ldots + p_k. \end{aligned}$$

Similarly

$$\begin{aligned} q_{n+1} + q_n + \ldots + q_{k+1} &= \alpha_{n+1}\beta_0 + \alpha_n\beta_1 + \ldots + \alpha_{k+1}\beta_{n-k} \\ &< \alpha_n\beta_0 + \alpha_{n-1}\beta_1 + \ldots + \alpha_k\beta_{n-k} \\ &= p_n + p_{n-1} + \ldots + p_k. \end{aligned}$$

In view of (5), this implies that

$$q_0 + q_1 + \ldots + q_k > p_0 + p_1 + \ldots + p_{k-1},$$

and the proof of (6) is complete.

For the second step in the proof of Theorem 1, we introduce symbols for the successive differences of the terms in (6). We put

$$r_0 = q_0,\; r'_0 = p_0 - q_0,\; r_1 = (q_0 + q_1) - p_0,\; r'_1 = (p_0 + p_1) - (q_0 + q_1), \ldots$$
$$r_n = (q_0 + \ldots + q_n) - (p_0 + \ldots + p_{n-1}),\; r'_n = q_{n+1}.$$

All these numbers are positive, and we have

$$p_0 = r_0 + r'_0,\; p_1 = r_1 + r'_1, \ldots, p_n = r_n + r'_n,$$
$$q_0 = r_0,\; q_1 = r'_0 + r_1, \ldots, q_n = r'_{n-1} + r_n,\; q_{n+1} = r'_n.$$

Hence, by (3) and (4),

$$w_n = r_0 a_0 b_n + r'_0 a_0 b_n + r_1 a_1 b_{n-1} + \ldots + r_n a_n b_0 + r'_n a_n b_0,$$
$$w_{n+1} = r_0 a_0 b_{n+1} + r'_0 a_1 b_n + r_1 a_1 b_n + \ldots + r_n a_n b_1 + r'_n a_{n+1} b_0.$$

These expressions render Theorem 1 immediate, on comparison of corresponding terms.

2. To prove Theorem 2, we use the following lemma:

LEMMA. *Let W_n be defined by*

$$W_n = a_0 b_n + \binom{n}{1} a_1 b_{n-1} + \binom{n}{2} a_2 b_{n-2} + \ldots + a_n b_0. \tag{7}$$

Then, if a_n and b_n are positive and logarithmically convex, so is W_n.

Proof. The desired result $W_n^2 \leq W_{n-1}W_{n+1}$ holds for $n = 1$ since

$$\begin{aligned} W_0W_2 - W_1^2 &= a_0b_0(a_0b_2 + 2a_1b_1 + a_2b_0) - (a_0b_1 + a_1b_0)^2 \\ &= a_0^2(b_0b_2 - b_1^2) + b_0^2(a_0a_2 - a_1^2) \geq 0. \end{aligned}$$

We prove it for general n by induction.

By the well-known property

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$$

of the binomial coefficients, we have, for $n \geq 1$,

$$W_n = W'_{n-1} + W''_{n-1},$$

where W'_{n-1} is formed with the sequences $a_1, a_2, \ldots$ and $b_0, b_1, \ldots$ and W''_{n-1} is formed with the sequences $a_0, a_1, \ldots$ and $b_1, b_2, \ldots$. By the hypothesis of the induction, applied to the two former sequences, we have

$$(W'_{n-1})^2 \leq W'_{n-2}W'_n,$$

and similarly

$$(W''_{n-1})^2 \leq W''_{n-2}W''_n.$$

By the inequality of the arithmetic and geometric means, it follows that

$$2\ W'_{n-1}W''_{n-1} \leq 2\ \left\{W'_{n-2}W'_nW''_{n-2}W''_n\right\}^{\frac{1}{2}} \leq W'_{n-2}W''_n + W''_{n-2}W'_n.$$

Hence, using again the hypothesis of the induction, we obtain

$$W_n^2 = (W'_{n-1} + W''_{n-1})^2 \leq W'_{n-2}W'_n + W'_{n-2}W''_n$$
$$+ W''_{n-2}W'_n + W''_{n-2}W''_n = W_{n-1}W_{n+1}.$$

This proves the Lemma.

An immediate corollary to the Lemma is that the same conclusion holds for $W_n(\lambda, \mu)$ defined by

$$W_n(\lambda, \mu) = a_0b_n\mu^n + \binom{n}{1}a_1\lambda b_{n-1}\mu^{n-1} + \ldots + a_n\lambda^n b_0, \tag{8}$$

where λ, μ are any two positive numbers.

We can now prove Theorem 2 as follows. By (1) and (2), we have

$$\alpha_m\beta_{n-m} = \binom{n}{m}\frac{\Gamma(\alpha+m)\Gamma(\beta+n-m)}{n!\,\Gamma(\alpha)\Gamma(\beta)}$$
$$= \binom{n}{m}\frac{1}{\Gamma(\alpha)\,\Gamma(\beta)}\int_0^1 t^{\alpha+m-1}(1-t)^{\beta+n-m-1}dt.$$

Substituting in (3), and using the notation of (8), we obtain

$$w_n = \frac{1}{\Gamma(\alpha)\,\Gamma(\beta)}\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}W_n(t, 1-t)\,dt.$$

Since $W_n(t, 1-t)$ is logarithmically convex for each t, it follows from the inequality of Schwarz that w_n is, since

$$\Gamma(\alpha)\Gamma(\beta)w_n \leq \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\left\{W_{n-1}(t, 1-t)W_{n+1}(t, 1-t)\right\}^{\frac{1}{2}}dt$$
$$\leq \left\{\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}W_{n-1}(t, 1-t)dt\right\}^{\frac{1}{2}}$$
$$\cdot\left\{\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}W_{n+1}(t, 1-t)dt\right\}^{\frac{1}{2}}$$
$$= (\Gamma(\alpha)\Gamma(\beta)w_{n-1}\Gamma(\alpha)\Gamma(\beta)w_{n+1})^{\frac{1}{2}}.$$

This proves Theorem 2.

3. The two theorems proved above have a certain resemblance to the following simple but useful theorem of Kaluza.[1]

If the a_n are positive and logarithmically convex, and

$$(a_0 + a_1x + a_2x^2 + \ldots)^{-1} = b_0 - b_1x - b_2x^2 - \ldots,$$

then all the b_n are positive.

All three theorems give conditions which ensure that a power series, derived from given power series by multiplication or division, shall have some simple property.

There is one class of power series to which our theorems can readily be applied. Suppose $\phi(t)$ is positive and integrable in the interval $(0, h)$, and let

$$\int_0^h \phi(t)(1 - xt)^{-\alpha}dt = \Sigma \alpha_n a_n x^n. \tag{9}$$

Then

$$a_n = \int_0^h \phi(t)t^n dt,$$

and the a_n, being the successive moments of a positive function, are logarithmically convex.

4. The particular problem from which our investigation started was that of showing that

$$\left[\int_0^1 (1 + u^4 - 2xu^2)^{-\frac{1}{2}}du\right]^{-2} + \left[\int_0^1 (1 + u^4 + 2xu^2)^{-\frac{1}{2}}du\right]^{-2} \tag{10}$$

decreases steadily as x increases from 0 to 1.

(It can be shown that the expression (10) represents $(2r_0\Lambda/\pi)^2$, where r_0 denotes the inner conformal radius of a rectangle with respect to its centre, and Λ denotes the principal frequency of vibration of a membrane with the rectangle as its boundary. The product $r_0\Lambda$ depends on the shape but not on the size of the rectangle, and the parameter x specifies this shape. As x increases from 0 to 1, the ratio of the two sides of the rectangle increases steadily from 1 to infinity. Our assertion concerning (10) means that the product $r_0\Lambda$ decreases steadily in this process.)

By the change of variable

$$2u^2/(1 + u^4) = t$$

the first integral in (10) is transformed into an integral $I(x)$ of the type (9), with $h = 1$ and $\alpha = 1/2$. Theorem 2, applied to this integral, tells us that the coefficients of the power series for $I^2(x)$ are logarithmically convex. From Kaluza's theorem, it follows that the expression (10) has the form

$$2b_0 - 2b_2x^2 - 2b_4x^4 - \ldots$$

with positive b_n. This obviously decreases as x increases.

We should perhaps observe that instead of using Theorem 2 in the above argument, we can use the following *ad hoc* argument. We have

$$I^2(x) = \int_0^1\int_0^1 \phi(t)\phi(t')(1 - xt)^{-\frac{1}{2}}(1 - xt')^{-\frac{1}{2}}\, dtdt'.$$

[1] *Math. Zeit.*, vol. 28 (1928), 161-170.

Let

$$(1 - xt)^{-\frac{1}{2}}(1 - xt')^{-\frac{1}{2}} = \Sigma\, A_n(t, t')x^n;$$

then

$$I^2(x) = \Sigma\, c_n x^n,$$

where

$$c_n = \int_0^1 \int_0^1 \phi(t)\phi(t')A_n(t, t')dtdt'.$$

If we prove that $A_n(t, t')$ is logarithmically convex, for fixed t, t', it will follow that c_n is logarithmically convex, as desired. In fact, it is easily seen that

$$A_n(t, t') = \frac{2}{\pi}\int_0^{\frac{\pi}{2}} (t \cos^2\theta + t' \sin^2\theta)^n d\theta,$$

and this is obviously logarithmically convex.

5. For the sake of completeness, we mention the integral analogues of Theorems 1 and 2, although they are less interesting.

Suppose that $f(x)$ and $g(x)$ are positive and integrable for $x \geq 0$, and bounded in any finite interval. We retain (1) and put

$$h(x) = \int_0^x t^{\alpha-1}f(t)(x - t)^{\beta-1}g(x - t)dt.$$

THEOREM 3. *If $f(x)$ and $g(x)$ are both monotonic increasing, so is $h(x)$, and if $f(x)$ and $g(x)$ are both monotonic decreasing, so is $h(x)$.*

THEOREM 4. *If $f(x)$ and $g(x)$ are both logarithmically convex, so is $h(x)$.*

We say that $f(x)$ is logarithmically convex, if for $x \geq d > 0$,

$$f^2(x) \leq f(x - d)f(x + d).$$

By changing the variable of integration and using (1), we obtain

$$h(x) = \int_0^1 u^{\alpha-1}(1 - u)^{\beta-1} f(ux)\, g((1 - u)x)\, du,$$

and this representation of $h(x)$ renders Theorem 3 obvious. By the hypothesis of Theorem 4 and Schwarz's inequality,

$$\begin{aligned} h(x) &\leq \int_0^1 u^{\alpha-1}(1 - u)^{\beta-1}\{f(u[x - d])\, f(u[x + d])\}^{\frac{1}{2}} \\ &\qquad \cdot \{g([1 - u][x - d])\, g([1 - u][x + d])\}^{\frac{1}{2}}\, du \\ &\leq \left\{\int_0^1 u^{\alpha-1}(1 - u)^{\beta-1} f(u[x - d])\, g([1 - u][x - d])du\right\}^{\frac{1}{2}} \\ &\qquad \cdot \left\{\int_0^1 u^{\alpha-1}(1 - u)^{\beta-1} f(u[x + d])\, g([1 - u][x + d])du\right\}^{\frac{1}{2}} \\ &= \{h(x - d)\, h(x + d)\}^{\frac{1}{2}}. \end{aligned}$$

This proves Theorem 4.

University College, London
Stanford University

SUR LES SURFACES A COURBURE MOYENNE ISOTHERME

VICTOR LALAN

Nous nous occupons dans ce travail d'une classe de surfaces dont l'équation différentielle est du cinquième ordre, et qui jouissent de la propriété que, sur elles, les lignes d'égale courbure moyenne forment, avec leurs trajectoires orthogonales, un système isotherme. Dans cette classe rentrent les surfaces admettant un groupe de déplacements à un paramètre (cylindres, surfaces de révolution et hélicoïdes), ainsi que les surfaces admettant une infinité de déformations avec conservation des courbures principales (surfaces d'Ossian Bonnet).

Notre méthode repose essentiellement sur l'emploi de certaines formes différentielles qui se sont présentées à nous dans l'étude des lignes minima des surfaces, mais qui sont susceptibles aussi d'une définition simple dans le domaine réel, comme nous le montrons au n° 1. On pourra consulter à ce sujet diverses Notes que nous avons communiquées à l'Académie des Sciences,[1] et aussi un mémoire qui doit paraître dans le Bulletin de la Société Mathématique de France de 1947. Nous supposons le lecteur initié aux méthodes développées par M. E. Cartan dans ses divers ouvrages.[2]

I. Définition et propriétés générales

1. Nous appelons *surfaces à courbure moyenne isotherme*, ou, plus brièvement, *surfaces HI*, les surfaces sur lesquelles les lignes d'égale courbure moyenne forment, conjointement avec leurs trajectoires orthogonales, un système isotherme. Leur étude, qui ne semble pas avoir été systématiquement entreprise, se trouve grandement facilitée par l'emploi de certaines formes différentielles, que nous avons appelées les *formes minima* de la surface, et qui ne sont autres que les différentielles des pseudo-arcs des lignes de longueur nulle de la surface. On peut d'ailleurs définir ces formes sans faire appel à la théorie des lignes minima, en opérant comme suit.

Si ds^2 et ϕ désignent les deux formes quadratiques de la surface, a et c les courbures principales, ϖ_1 et ϖ_2 les arcs élémentaires des lignes de courbure, on a

$$ds^2 = \varpi_1^2 + \varpi_2^2, \quad \phi = a\varpi_1^2 + c\varpi_2^2,$$

donc

$$\phi - c\,ds^2 = (a - c)\varpi_1^2, \quad \phi - a\,ds^2 = (c - a)\varpi_2^2.$$

Received February 11, 1948.

[1] *Comptes rendus*, 1946, 1947, 1948, *passim*.

[2] *La théorie des groupes finis et la géométrie différentielle* (Paris, 1937). *Les systèmes différentiels extérieurs et leurs applications géométriques* (Paris, 1945). Voir aussi, sur les surfaces d'O. Bonnet, un mémoire de M. E. Cartan paru dans le *Bull. des Sciences Math.*, vol. 66 (1942), 55-85, où l'on trouvera des indications bibliographiques.

Supposons $a > c$, ce qui est loisible en tout point qui n'est pas un ombilic, et posons

(1) $$\theta_1 = \sqrt{\phi - c ds^2} = \sqrt{a-c}\,\varpi_1, \quad \theta_2 = \sqrt{a ds^2 - \phi} = \sqrt{a-c}\,\varpi_2.$$

Les formes minima, ω_1 et ω_2, se définissent à partir de là

(2) $$2\omega_1 = \theta_1 - i\theta_2, \quad 2\omega_2 = \theta_1 + i\theta_2.$$

Il est donc parfaitement équivalent d'utiliser les formes réelles θ_1, θ_2, que nous appellerons les *formes principales*, ou les formes imaginaires conjuguées ω_1, ω_2. En revanche, nombre de propriétés apparaissent quand on emploie θ_1 et θ_2 (ou ω_1 et ω_2), qui restent cachées tant qu'on s'en tient aux formes de Darboux-Cartan ϖ_1 et ϖ_2.

Nous poserons

$$H = \frac{a+c}{2}, \quad A = \frac{a-c}{2}.$$

H est la courbure moyenne, A sera appelée l'*asphéricité*; on suppose la région étudiée dépourvue de singularités et d'ombilics, et l'on s'arrange pour que A soit positive.

Les formes quadratiques de la surface s'écrivent, comme on le voit immédiatement,

(3) $$ds^2 = \frac{2\omega_1\omega_2}{A}, \quad \phi = \omega_1^2 + 2\frac{H}{A}\omega_1\omega_2 + \omega_2^2.$$

Définissons en outre les invariants minima r et s par

(4) $$d\omega_1 = r[\omega_1\omega_2], \quad d\omega_2 = s[\omega_2\omega_1]$$

et les invariants principaux ρ et σ par

(5) $$d\theta_1 = \rho[\theta_1\theta_2], \quad d\theta_2 = \sigma[\theta_2\theta_1].$$

On vérifie sans peine que

(6) $$r = \sigma - i\rho, \quad s = \sigma + i\rho.$$

2. Le premier résultat important que nous obtenons grâce à l'emploi des formes principales, c'est l'expression des équations classiques de Codazzi sous la forme d'une équation unique aux différentielles totales.

Dans la méthode du trièdre mobile de M. E. Cartan, on a

$$\varpi_{12} = h\varpi_1 + k\varpi_2$$

avec

(7) $$d\varpi_1 = h[\varpi_1\varpi_2], \quad d\varpi_2 = k[\varpi_1\varpi_2]$$

et

$$\varpi_{13} = a\varpi_1, \quad \varpi_{23} = c\varpi_2.$$

Or, deux des conditions d'intégrabilité s'écrivent

(8) $$a_2 = h(a-c), \quad c_1 = k(a-c).$$

Dans ces formules, les indices 1 et 2 qui affectent c et a désignent des dérivées relatives à ϖ_1, ϖ_2, suivant le schéma

$$df = f_1\varpi_1 + f_2\varpi_2.$$

Introduisons pareillement les dérivées T_1f, T_2f relatives à θ_1, θ_2, vérifiant

$$df = T_1f\theta_1 + T_2f\theta_2,$$

de sorte que, compte tenu de (1),

$$T_1f = \frac{f_1}{\sqrt{a-c}}, \quad T_2f = \frac{f_2}{\sqrt{a-c}}. \tag{9}$$

La différentiation extérieure de (1) donne

$$\rho = +\frac{1}{a-c}\left(\frac{h\sqrt{a-c}}{2} + \frac{c_2}{2\sqrt{a-c}}\right), \tag{10}$$

$$\sigma = -\frac{1}{a-c}\left(\frac{k\sqrt{a-c}}{2} + \frac{a_1}{2\sqrt{a-c}}\right),$$

d'où, en résolvant,

$$a_1 = -k(a-c) - 2(a-c)^{\frac{3}{2}}\sigma, \quad c_2 = -h(a-c) + 2(a-c)^{\frac{3}{2}}\rho.$$

Par conséquent, en tenant compte de (8),

$$T_1(a+c) = -2(a-c)\sigma, \quad T_2(a+c) = 2(a-c)\rho, \tag{11}$$

et, puisque $a + c = 2H$, les deux équations de Codazzi se condensent, comme annoncé, en une équation aux différentielles totales,

$$dH = 2A(-\sigma\theta_1 + \rho\theta_2), \tag{12}$$

que nous utiliserons plutôt sous la forme équivalente,

$$dH = -2A(r\omega_1 + s\omega_2). \tag{13}$$

3. Supposons que les lignes $H = C$ forment une famille de courbes isothermes sur la surface. Il existe, en conséquence, des paramètres isotropes u, v, respectivement intégrales premières de ω_1, ω_2, tels que l'on ait $H = f(u+v)$. Nous prenons u et v imaginaires conjuguées, si bien que $u + v$ est une fonction harmonique réelle sur la surface. La variable complexe v sera dite *attachée* aux courbes $H = C$, en ce sens que les courbes $H = C$ s'obtiennent en égalant à une constante la partie réelle de v. Puisque $H = f(u + v)$, l'équation $dH = 0$, équivalente d'après (13) à $r\omega_1 + s\omega_2 = 0$, doit être équivalente à $du + dv = 0$. Or, posons $\omega_1 = \alpha(u, v)du$, $\omega_2 = \beta(u, v)dv$, nous aurons, d'après (4),

$$r = -\frac{\alpha_v}{\alpha\beta}, \quad s = -\frac{\beta_u}{\alpha\beta}$$

et

$$r\omega_1 + s\omega_2 = -\frac{\alpha_v}{\beta}du - \frac{\beta_u}{\alpha}dv;$$

cette dernière expression ne contiendra $du + dv$ en facteur que si

$$\frac{\alpha_v}{\beta} = \frac{\beta_u}{\alpha}, \text{ ou } \alpha\alpha_v = \beta\beta_u,$$

c'est-à-dire, si la forme $\alpha^2du + \beta^2dv$ est une différentielle exacte $d\psi$, si, par conséquent, $\alpha = \sqrt{\psi_u}$, $\beta = \sqrt{\psi_v}$, et

(14) $$\omega_1 = \sqrt{\psi_u}\, du\,, \quad \omega_2 = \sqrt{\psi_v}\, dv\,,$$
d'où cette proposition: *si les lignes d'égale courbure moyenne d'une surface forment une famille de courbes isothermes, il existe des coordonnées isotropes u, v, et une fonction $\psi(u, v)$ telles que les formes minima s'écrivent $\omega_1 = \sqrt{\psi_u}\, du$, $\omega_2 = \sqrt{\psi_v}\, dv$, tandis que la courbure moyenne est $H = f(u + v)$.*

Nous appellerons $\psi(u, v)$ la *fonction primitive* de la surface HI, pour rappeler que les formes minima d'une telle surface se déduisent de ψ par des dérivations.

4. Désignons par ϕ_1 la forme $r\omega_1 + s\omega_2$, et par ϕ_2, $i(r\omega_1 - s\omega_2)$, les lignes $\phi_1 = 0$, $\phi_2 = 0$ sont orthogonales, car

(15) $$\phi_1^2 + \phi_2^2 = (2Ars)\left(\frac{2\omega_1\omega_2}{A}\right),$$

et le second facteur du second membre est l'élément linéaire de la surface (formule (3)). Posons ensuite

(16) $$d\phi_1 = R\,[\phi_1\,\phi_2]\,, \qquad d\phi_2 = S\,[\phi_2\,\phi_1]\,.$$

La forme

(17) $$\chi = S\phi_1 + R\phi_2$$

est importante à considérer, comme nous allons le voir; elle s'écrit, en tenant compte des formules

(18) $$S = \frac{r^2 + s^2 - s_1 - r_2}{2rs}\,, \quad R = i\,\frac{r^2 - s^2 + s_1 - r_2}{2rs}\,,$$

qui s'obtiennent par différentiation extérieure,

(19) $$\chi = \left(s - \frac{s_1}{s}\right)\omega_1 + \left(r - \frac{r_2}{r}\right)\omega_2.$$

(Dans les deux formules précédentes, les indices 1 et 2 affectant r et s désignent des dérivées relatives à ω_1 et ω_2; il en sera de même par la suite.)

Si $H = f(u + v)$, ce qui entraîne (14), on trouve que χ est une différentielle exacte, à savoir

(20) $$\chi = d \log \frac{\sqrt{\psi_u\psi_v}}{\psi_{uv}}\,.$$

Réciproquement, si χ est une différentielle exacte, $e^{\int\chi}$ est facteur intégrant à la fois pour ϕ_1 et ϕ_2, car on vérifie sans peine que

$$d(e^{\int\chi}\phi_1) = 0, \quad d(e^{\int\chi}\phi_2) = 0.$$

On peut donc poser

$$e^{\int\chi}\phi_1 = dp, \quad e^{\int\chi}\phi_2 = dq,$$

et, par suite, l'élément linéaire de la surface devient, d'après (15),

$$\frac{e^{-2\int\chi}}{2Ars}\,(dp^2 + dq^2)\,,$$

ce qui montre que les lignes $\phi_1 = 0$, $\phi_2 = 0$, c'est-à-dire les lignes $H = C$ et leurs trajectoires orthogonales, forment un système isotherme. D'où la proposition: *la condition nécessaire et suffisante pour qu'une surface soit à courbure moyenne isotherme* (*surface HI*), *c'est que la forme χ, de la formule* (19), *soit une différentielle exacte.*

Les formules spéciales aux surfaces *HI* sont résumées ci-dessous:

$$(21)\quad \begin{cases} \omega_1 = \sqrt{\psi_u}\, du, \qquad \omega_2 = \sqrt{\psi_v}\, dv, \\ r = -\dfrac{\psi_{uv}}{2\psi_u\sqrt{\psi_v}}, \qquad s = -\dfrac{\psi_{uv}}{2\psi_v\sqrt{\psi_u}}, \\ r\omega_1 + s\omega_2 = \phi_1 = -\dfrac{\psi_{uv}}{2\sqrt{\psi_u\psi_v}}(du+dv),\ \chi = d\log\dfrac{\sqrt{\psi_u\psi_v}}{\psi_{uv}}, \\ H = f(u+v), \qquad A = \dfrac{\sqrt{\psi_u\psi_v}}{\psi_{uv}} f'(u+v). \end{cases}$$

5. Les deux formes quadratiques d'une surface *HI* s'écrivent

$$(22)\qquad ds^2 = 2\frac{\psi_{uv}}{f'(u+v)}(du+dv),\ \phi = \psi_u du^2 + 2f\frac{\psi_{uv}}{f'}dudv + \psi_v dv^2$$

Elles ne contiennent que les deux fonctions $\psi(u, v)$ et $f(u+v)$. Ces deux fonctions sont d'ailleurs liées par la 3[e] condition d'intégrabilité, qui traduit le *theorema egregium* de Gauss. Quand on écrit les formes

$$ds^2 = 2\ F\ dudv,\ \phi = Ldu^2 + 2\ Mdudv + Ndv^2,$$

ce théorème s'exprime par

$$\mathbf{K} = -\frac{1}{F}\frac{\partial^2}{\partial u \partial v}(\log F) \qquad (\mathbf{K} = H^2 - A^2 \text{ courbure totale}).$$

Cette formule devient ici, compte tenu des expressions de H, A et F,

$$(23)\qquad \frac{d^2}{dt^2}\log|f'(t)| = \frac{\partial^2}{\partial u\partial v}\log|\psi_{uv}| + \psi_{uv}\frac{f^2}{f'} - \frac{\psi_u\psi_v}{\psi_{uv}}f' \qquad (t = u+v).$$

On remarquera que, sur une surface réelle, ψ_{uv} et $f'(u+v)$ sont de même signe, d'après la première équation (22), mais ce signe peut être quelconque.

Dans ce qui précède, nous avons supposé implicitement que $\psi(u, v)$ n'était pas une fonction harmonique sur la surface; au cas contraire, on aurait $\psi = U + V$, $\omega_1 = \sqrt{U'}du$, $\omega_2 = \sqrt{V'}dv$, ω_1 et ω_2 seraient des différentielles exactes, d'où, d'après (4), r et s seraient nuls, et, d'après (13), H serait constant. Quand nous parlerons de surfaces *HI*, nous supposerons toujours que H n'est pas constant, et, partant, que ψ n'est pas une fonction harmonique sur la surface.

6. La fonction primitive $\psi(u, v)$ n'est pas invariante, car la fonction harmonique $t = u + v$, qui intervient dans sa définition, n'est pas définie univoquement; elle n'est assujettie qu'à la condition d'être constante sur les lignes $H = C$. On peut la remplacer par $\bar{t} = at + b$ (a, b, constantes). Dans ce changement, $\psi_u du^2$, qui est ω_1^2, devient $a\psi_{\bar{u}}\cdot\dfrac{d\bar{u}^2}{a^2} = \psi_{\bar{u}}\dfrac{d\bar{u}^2}{a}$, ce qui peut bien s'écrire $\bar{\psi}_{\bar{u}}d\bar{u}^2$, mais à condition de poser $\bar{\psi} = \dfrac{\psi}{a}$. Ainsi, quand on remplace t par $at + b$, ψ doit être remplacé par $\dfrac{\psi}{a}$, de sorte que ψdt reste invariant.

Les *courbes primitives*, $\psi(u, v) = C$, de la surface HI, sont dans une relation remarquable avec les courbes $H = C$ et les lignes de courbure. De leur équation différentielle, $\psi_u du + \psi_v dv = 0$, qui s'écrit aussi $\sqrt{\psi_u}\omega_1 + \sqrt{\psi_v}\omega_2 = 0$, on déduit qu'elles coupent les premières lignes de courbure sous un angle α donné par

$$e^{2i\alpha} = -\frac{\sqrt{\psi_u}}{\sqrt{\psi_v}};$$

or, les lignes $H = C$ ont pour équation $r\omega_1 + s\omega_2 = 0$, c'est-à-dire, $\sqrt{\psi_v}du + \sqrt{\psi_u}dv = 0$; elles coupent donc les premières lignes de courbure sous un angle β qui vérifie

$$e^{2i\beta} = -\frac{\sqrt{\psi_v}}{\sqrt{\psi_u}} = e^{-2i\alpha},$$

donc $\beta = -\alpha$, et nous avons la proposition: *sur une surface HI, les lignes primitives et les lignes d'égale courbure moyenne sont bissectées par les lignes de courbure.*

7. La proposition énoncée au n° 4 nous permet d'obtenir sous forme invariante l'équation différentielle des surfaces HI. La condition que la forme χ soit une différentielle exacte s'exprime en effet par une relation du 5^e ordre, qui n'est autre que l'équation cherchée. Elle s'écrit

$$r_1 - [(\log r)_{21} - s(\log r)_2] = s_2 - [(\log s)_{12} - r(\log s)_1].$$

Introduisons le second paramètre différentiel de Beltrami, dont l'expression, pour une fonction f quelconque, est

$$\Delta_2 f = 2A(f_{21} - sf_2), \text{ et aussi, } 2A(f_{12} - rf_1).$$

L'équation précédente devient

$$r_1 - \frac{\Delta_2(\log r)}{2A} = s_2 - \frac{\Delta_2(\log s)}{2A}$$

ou enfin

$$\Delta_2\left(\log\frac{r}{s}\right) = 2A(r_1 - s_2), \tag{24}$$

qui est, en définitive, l'équation des surfaces HI. On peut du reste la formuler autrement.

β étant toujours l'angle sous lequel les courbes $H = C$ coupent les premières lignes de courbure, on a

$$e^{2i\beta} = -\frac{r}{s},$$

donc

$$\Delta_2\left(\log\frac{r}{s}\right) = 2i\Delta_2\beta.$$

Par ailleurs, la forme $s\omega_1+r\omega_2$ a pour différentielle extérieure $(r_1-s_2)[\omega_1\omega_2]$, donc

$$r_1 - s_2 = \frac{d(s\omega_1 + r\omega_2)}{[\omega_1\omega_2]}$$

et l'équation (24) devient

$$i\Delta_2\beta = A\,\frac{d(s\omega_1 + r\omega_2)}{[\omega_1\omega_2]}. \tag{25}$$

Nous retrouverons la forme $s\omega_1+r\omega_2$ au paragraphe suivant.

II. Surfaces HI isothermiques

8. Une surface est isothermique si les arcs élémentaires des lignes de courbure ont un facteur intégrant commun. D'après (1) et (2), les formes minima en auront un aussi. Appelons-le μ, et exprimons que $\mu\omega_1$ et $\mu\omega_2$ sont des différentielles exactes; il vient

$$\mu_2 - r\mu = 0, \quad \mu_1 - s\mu = 0$$

d'où

$$\frac{d\mu}{\mu} = s\omega_1 + r\omega_2,$$

ce qui s'énonce: *sur toute surface isothermique, la forme $s\omega_1+r\omega_2$ est une différentielle exacte, et $e^{\int s\omega_1+r\omega_2}$ est un facteur intégrant à la fois pour ω_1 et pour ω_2.*

Pour rappeler ce rôle de $s\omega_1+r\omega_2$, nous l'appellerons la *forme isothermique* de la surface. L'équation différentielle des surfaces isothermiques se déduit de ce qui précède; en exprimant que $s\omega_1+r\omega_2$ est une différentielle exacte, on trouve l'équation

$$r_1 - s_2 = 0\,;$$

elle est du quatrième ordre.

Les lignes de courbure des surfaces isothermiques forment un réseau isotherme; on peut donc leur attacher une variable complexe z. Pour cela, posons

$$e^{\int s\omega_1+r\omega_2}\omega_1 = dz_0, \quad e^{\int s\omega_1+r\omega_2}\omega_2 = dz \tag{26}$$

$$(z_0 = x - iy,\ z = x + iy).$$

La variable complexe z répond à la question, car on voit, d'après (1) et (2), que

$$dx = \sqrt{\frac{A}{2}}\,e^{\int s\omega_1+r\omega_2}\varpi_1, \quad dy = \sqrt{\frac{A}{2}}\,e^{\int s\omega_1+r\omega_2}\varpi_2, \tag{27}$$

ce qui montre qu'on obtient bien les premières lignes de courbure, $\varpi_2=0$, en égalant à une constante la partie imaginaire de z, et les secondes, la partie réelle.

9. Soit maintenant une surface à la fois HI et isothermique. La formule (25) montre qu' alors, l'angle β sous lequel les lignes $H = C$ coupent les premières lignes de courbure est une fonction harmonique, ce qui s'explique, puisque les lignes $H = C$ sont isothermes. Il s'ensuit que, dans ce cas, les courbes primitives $\psi(u, v) = C$ sont isothermes, elles aussi, car elles coupent les premières lignes de courbure sous un angle α qui, d'après le n° **6**, vaut $-\beta$, et,

par conséquent, est harmonique. ψ est donc une fonction de fonction harmonique: $\psi = g(U + V)$.

Ce résultat se vérifie facilement par le calcul. En effet, la forme isothermique, sur une surface HI, s'écrit d'après (21),

$$s\omega_1 + r\omega_2 = -\frac{1}{2}\frac{\psi_{uv}}{\psi_u\psi_v}d\psi. \tag{28}$$

Pour que la surface soit isothermique, il faut donc, et il suffit, que $\dfrac{\psi_{uv}}{\psi_u\psi_v}$, ou $\dfrac{\Delta_2\psi}{\Delta_1\psi}$, soit fonction de ψ. Or cela signifie précisément que les courbes $\psi = C$ sont isothermes, que ψ, par conséquent, est de la forme $g(U + V)$. Donc, *la condition nécessaire et suffisante pour qu'une surface HI soit isothermique, c'est que la fonction primitive ψ puisse s'écrire $g(U + V)$.*

U et V sont imaginaires conjuguées. On peut regarder V comme la variable complexe attachée au réseau isotherme formé des courbes primitives et de leurs trajectoires orthogonales.

Les trois variables complexes v, z, et V sont trois intégrales premières de ω_2, elles sont donc fonction l'une de l'autre. En particulier, V est fonction de $x + iy$:

$$V = P(x, y) + iQ(x, y) \qquad (\text{et } U = P(x, y) - iQ(x, y))$$

de telle sorte que

$$U + V = 2P \quad \text{et} \quad \psi = g(2P).$$

Pour une surface HI isothermique, les formules (26) se simplifient. Exprimons d'abord le facteur intégrant $e^{\int s\omega_1 + r\omega_2}$. On a vu que

$$s\omega_1 + r\omega_2 = -\frac{1}{2}\frac{\psi_{uv}}{\psi_u\psi_v}d\psi.$$

Or, puisque $\psi = g(U + V)$, on a, inversement, $U + V = G(\psi)$, et

$$\frac{\psi_{uv}}{\psi_u\psi_v} = \frac{g''}{g'^2} = -\frac{G''(\psi)}{G'(\psi)};$$

donc, à un facteur constant près,

$$e^{\int s\omega_1 + r\omega_2} = \sqrt{G'(\psi)}$$

et, par suite,

$$e^{\int s\omega_1 + r\omega_2}\omega_2 = \sqrt{G'(\psi)} \cdot \sqrt{\psi_v}\, dv = \sqrt{G_v}\, dv = \sqrt{V'}\, dv.$$

Les formules (26) deviennent donc

$$dz_0 = \sqrt{U'}\, du, \quad dz = \sqrt{V'}\, dv. \tag{29}$$

Cette dernière formule peut s'écrire

$$dz^2 = dVdv,$$

ce qui met en évidence le fait, énoncé au n° 6, que les lignes de courbure bissectent les lignes $H = C$ et $\psi = C$.

III. Surfaces *HI* isothermiques et *W*

10. Si la surface est W, c'est-à-dire, s'il existe une relation entre H et A, la formule (13) montre que $r\omega_1 + s\omega_2$ est une différentielle exacte, et réciproquement. De là découle l'équation différentielle des surfaces W: en exprimant que $r\omega_1 + s\omega_2$ a une différentielle extérieure nulle, on trouve l'équation du 4ᵉ ordre

$$s_1 - r_2 + r^2 - s^2 = 0,$$

qui peut s'écrire aussi

$$\frac{s - \dfrac{s_1}{s}}{r} = \frac{r - \dfrac{r_2}{r}}{s}\,;$$

sur les surfaces W, χ et ϕ_1 ne sont donc pas indépendantes; c'est ce qui ressort de la relation (n° 4)

$$d\phi_1 = [\phi_1\chi]\,.$$

11. Si une surface est à la fois W et isothermique, les formes $r\omega_1 + s\omega_2$ et $s\omega_1 + r\omega_2$ sont, l'une et l'autre, des différentielles exactes. Posons donc

$$s\omega_1 + r\omega_2 = \frac{d\lambda}{\lambda}\,, \quad r\omega_1 + s\omega_2 = \frac{d\mu}{\mu}\,.$$

Par addition, puis par soustraction, il vient

$$(s + r)(\omega_1 + \omega_2) = d\log(\lambda\mu), \quad (s - r)(\omega_1 - \omega_2) = d\log\frac{\lambda}{\mu}\,.$$

Or, $\omega_1 + \omega_2$ est proportionnel à ϖ_1, lui-même proportionnel à dx, d'après (27). Donc $\lambda\mu$ ne dépend que de x et, de même, $\dfrac{\lambda}{\mu}$ ne dépend que de y. Par conséquent

$$\lambda = X(x)Y(y), \qquad \mu = \frac{X(x)}{Y(y)}$$

d'où cette proposition: *sur toute surface W isothermique, la forme isothermique est la différentielle logarithmique d'une fonction $X(x)Y(y)$, et la forme $\phi_1 = r\omega_1 + s\omega_2$ est la différentielle logarithmique d'une fonction $X(x)/Y(y)$, x et y étant les variables harmoniques associées qui restent constantes le long des lignes de courbure.*

12. Passons à l'examen du cas où la surface serait à la fois HI, isothermique, et W. Nous avons obtenu le résultat suivant, que nous croyons nouveau: *les seules surfaces qui soient à la fois HI, isothermiques, et W sont, outre les surfaces de révolution, les cylindres, et certains cônes, celles sur lesquelles la fonction primitive ψ est de la forme $k\log(U+V)$, U et V étant telles que $\dfrac{U'V'}{(U+V)^2}$ ne dépende que de $u+v$.* Ces dernières surfaces, nous le verrons par la suite, sont les surfaces d'Ossian Bonnet de troisième classe.

La condition que la surface HI soit isothermique se traduit par $\psi = g(U+V)$ (n° **9**); celle que cette surface soit W, si l'on tient compte de l'expression de A (formule 21), s'exprime par

$$\frac{\psi_{uv}}{\sqrt{\psi_u \psi_v}} = \rho(u+v).$$

En combinant ces deux conditions, on obtient

$$\sqrt{U'V'} = \rho(u+v) \cdot \frac{g'(U+V)}{g''(U+V)}. \tag{30}$$

On satisfait à cette équation en supposant que $U + V$ est fonction de $u + v$, une fonction linéaire naturellement. A cause de l'indétermination qui subsiste dans la définition de u et v (n° **6**), on peut alors prendre simplement $U = u$ $V = v$. Les formes quadratiques d'une telle surface seraient

$$ds^2 = 2\frac{g''(u+v)}{f'(u+v)} du dv,$$

$$\phi = g'(u+v)du^2 + 2f(u+v)\frac{g''(u+v)}{f'(u+v)} du dv + g'(u+v)dv^2.$$

Tous les coefficients sont fonction de $u + v$. Les surfaces correspondantes admettent donc une infinité de déplacements sur elles-mêmes, par $u' = u + ia$, $v' = v - ia$, et une infinité de symétries par $u' = v + ib$, $v' = u - ib$. Les lignes $u + v =$ const., qui glissent sur elles-mêmes et admettent des symétries par rapport à des plans, ne peuvent être que des droites ou des cercles. Si ce sont des cercles, on a des *surfaces de révolution*, c'est le cas général. Si ce sont des droites, c'est-à-dire si $g' = Cf$, on a des *cylindres*. Nous n'insistons pas davantage sur ce cas simple.

13. Pour écarter la solution précédente, supposons que $U + V = \tau$ et $u + v = t$ soient des fonctions *indépendantes*. On doit déterminer $g(\tau)$ et les fonctions $U(u)$ et $V(v)$ pour que (30) soit satisfaite, mais il faut en outre que l'équation de Gauss (23) soit vérifiée; celle-ci s'écrit, en appelant $\mathbf{K}$ la courbure totale,

$$\frac{d^2}{dt^2} \log |f'| = \frac{\partial^2}{\partial u \partial v} \log |\psi_{uv}| + \frac{\psi_{uv}}{f'} \mathbf{K}. \tag{31}$$

La courbure totale $\mathbf{K}$ est, ici, fonction de t. Par ailleurs, $\psi = g(\tau)$ donne $\psi_{uv} = g''U'V'$, si bien que

$$\frac{\partial^2}{\partial u \partial v} \log |\psi_{uv}| = \frac{\partial^2}{\partial u \partial v} \log |g''| = \frac{d^2}{d\tau^2} \log |g''| \,.\, U'V',$$

ou, en tenant compte de (30)

$$\frac{\partial^2}{\partial u \partial v} \log |\psi_{uv}| = \rho^2(t) \cdot \frac{g'^2}{g''^2} (\log |g''|)''.$$

L'équation (31) devient donc, en utilisant de nouveau (30) pour exprimer ψ_{uv},

$$\frac{d^2}{dt^2} \log |f'| = \rho^2 \frac{g'^2}{g''^2} (\log |g''|)'' + \frac{g'^2}{g''} \cdot \frac{\rho^2}{f'} \mathbf{K}. \tag{32}$$

Dérivons par rapport à τ qui, par hypothèse, est indépendant de t:

$$0 = \left\{\frac{g'^2}{g''^2}(\log|g''|)''\right\}'_\tau + \frac{\mathbf{K}(t)}{f'(t)}\left(\frac{g'^2}{g''}\right)'_\tau. \tag{33}$$

Nous avons divisé par ρ^2 qui ne peut être nul, d'après (30), sans que g″ le soit, c'est-à-dire, sans que ψ soit harmonique, ce qui est exclu.

A. L'équation (33) est satisfaite si $\left(\dfrac{g'^2}{g''}\right)'_\tau = 0$; cela donne, en effet, $\dfrac{g'^2}{g''} = \dfrac{1}{m}$, d'où, comme le montre un calcul facile, $\dfrac{g'^2}{g''^2}(\log|g''|)'' = 2.$ Cette solution s'écrit aussi $\dfrac{\psi_{uv}}{\psi_u\psi_v} = m$: c'est une condition qui, nous le verrons plus loin, caractérise les surfaces d'Ossian Bonnet. En l'intégrant, on trouve

$$\psi = -\frac{1}{m}\log(U + V).$$

L'équation (30) donne alors

$$\frac{\sqrt{U'V'}}{U+V} = -\rho(u+v),$$

et nous montrerons (n° **24**) que cette condition détermine U, V, et ρ. L'équation de Gauss (32) devient

$$\frac{d^2}{dt^2}\log|f'| = 2\rho^2 + \frac{1}{m}\frac{\rho^2}{f'}\left(f^2 - \frac{f'^2}{\rho^2}\right).$$

Quand ψ, (c'est-à-dire U et V), a été déterminé, et ρ en conséquence, f n'est assujettie qu'à vérifier cette équation différentielle du troisième ordre: il y a donc une triple infinité de surfaces essentiellement différentes correspondant à la même fonction primitive ψ.

14. B. Cherchons à satisfaire (33) autrement, en supposant $\left(\dfrac{g'^2}{g''}\right) \neq 0$; elle peut alors s'écrire

$$\frac{\left\{\dfrac{g'^2}{g''^2}(\log|g''|)''\right\}'_\tau}{\left(\dfrac{g'^2}{g''}\right)'_\tau} = -\frac{\mathbf{K}(t)}{f'(t)}. \tag{34}$$

La valeur commune de ces rapports ne peut être qu'une constante, soit a, d'où en intégrant ce qui concerne g,

$$\frac{g'^2}{g''^2}(\log|g''|)'' = a\frac{g'^2}{g''} + b. \tag{35}$$

Mais, de l'équation (30), on peut déduire une autre équation différentielle

que doit vérifier g. Prenons le logarithme des deux membres, et dérivons d'abord par rapport à u, puis par rapport à v, il vient

$$0 = (\log \rho)''_{t^2} + \left(\log \frac{g'}{g''}\right)''_{\tau^2} U'V',$$

ou, en remplaçant $U'V'$ au moyen de (30), et séparant les variables,

$$\frac{\left(\log \frac{g''}{g'}\right)''_{\tau^2}}{\frac{g''^2}{g'^2}} = \frac{(\log \rho)''_{t^2}}{\rho^2}. \tag{36}$$

Comme précédemment, la valeur commune de ces rapports est une constante, c, d'où, pour g,

$$\left(\log \frac{g''}{g'}\right)'' = c \frac{g''^2}{g'^2}. \tag{37}$$

Il faut chercher les solutions communes à (35) et (37). Eliminant $(\log |g''|)''$ entre ces deux équations, il vient

$$(\log |g'|)'' = ag'' + (b - c)\frac{g''^2}{g'^2}, \tag{38}$$

qui peut s'écrire

$$\frac{g'''}{g''} = ag' + (b - c + 1)\frac{g''}{g'},$$

ou encore

$$(\log |g''|)' = ag' + (b - c + 1)(\log |g'|)'.$$

Dérivons en tenant compte de (35) et de (38):

$$ag'' + b\frac{g''^2}{g'^2} = ag'' + (b - c + 1)\left[ag'' + (b - c)\frac{g''^2}{g'^2}\right],$$

et, en divisant par g'', qui n'est pas nul, puisque ψ n'est pas harmonique,

$$[c - (b - c)^2]\frac{g''}{g'^2} = a(b - c + 1). \tag{39}$$

Or notre hypothèse actuelle est que $\frac{g''}{g'^2}$ n'est pas une constante; il faut donc que (39) s'évanouisse, c'est-à-dire qu'on ait

$$c = (b - c)^2 \text{ et } a(b - c + 1) = 0,$$

d'où deux hypothèses possibles:

$$(a) \qquad c = (b - c)^2, \quad b - c + 1 = 0,$$

$$(b) \qquad c = (b - c)^2, \quad a = 0.$$

15. L'hypothèse (a) équivaut à $c = 1, b = 0$. L'équation (35) donne alors $(\log |g''|)'' = ag''$, d'où $\log |g''| = ag + p\tau + q$. L'équation (37) devient

$(\log|g''|)'' = (\log|g'|)'' + \frac{g''^2}{g'^2}$. Portons-y la valeur trouvée pour $\log|g''|$, et développons le second membre:

$$ag'' = \frac{g'''}{g'}, \text{ d'où } \log|g''| = ag + q,$$

ce qui est compatible avec l'expression antérieure, en y faisant $p = 0$. Ainsi g peut être déterminée de façon à satisfaire (35) et (37).

Cherchons maintenant à déterminer $f(t)$. L'équation (34) donne

$$\mathbf{K} = -af'.$$

Or $\mathbf{K} = H^2 - A^2 = f^2 - \frac{f'^2}{\rho^2}$; l'équation ci-dessus donne donc $\rho^2 = \frac{f'^2}{f^2 + af'}$.

Mais on a, d'après (36), $(\log\rho)'' = \rho^2$, ce qui, exprimé en f, devient

$$(\log f'^2)'' - [\log(f^2 + af')]'' = \frac{2f'^2}{f^2 + af'}. \tag{40}$$

Par ailleurs, f doit satisfaire à l'équation (32), qui, compte tenu de l'expression trouvée pour $\log|g''|$, se réduit à

$$\frac{d^2}{dt^2}\log|f'| = 0. \tag{41}$$

Les équations (40) et (41) n'ont aucune solution commune, comme on s'en assure sans peine: l'hypothèse (a) ne donne donc rien.

16. Dans l'hypothèse (b), on a

$$c = (b - c)^2, \quad a = 0.$$

De $a = 0$, on déduit que, dans (33), le premier terme du second membre est nul (d'après (35)); il faut donc que le second soit nul aussi, donc $\mathbf{K} = 0$, les surfaces seront *développables*. Ce seront des développables sans arête de rebroussement, puisqu'elles doivent être isothermiques; ce ne seront pas des cylindres, pour lesquels τ et t ne seraient pas indépendants (n° **12**): ce seront donc des *cônes*. L'équation (35) se réduit à

$$\frac{g'^2}{g''^2}(\log|g''|)'' = b;$$

on calcule c par (37). Portant ces expressions de b et c dans $c = (b - c)^2$, on trouve, après simplification,

$$\frac{g''^2}{g'^2}\left(\log\frac{g''}{g'}\right)'' = [(\log g')'']^2,$$

d'où l'on tire $g' = (p\tau + q)^\alpha$, ou plus simplement, puisque τ n'est défini qu'à une transformation linéaire près,

$$g' = \tau^\alpha. \quad \left(b = \frac{1-\alpha}{\alpha^2}, \; c = \frac{1}{\alpha^2}\right).$$

On doit écarter $\alpha = -1$, car cela entraînerait $\frac{g''}{g'^2} = \text{const.}$, ce qui est exclu.

Donc $g = \dfrac{1}{\alpha+1}\tau^{\alpha+1}$, c'est-à-dire, $\psi = \dfrac{1}{\alpha+1}(U+V)^{\alpha+1}$.

Voyons si l'on peut déterminer $f(t)$. Puisque $\mathbf{K} = 0$, on a $f^2 - \dfrac{f'^2}{\rho^2} = 0$, donc $\rho^2 = \dfrac{f'^2}{f^2}$. L'équation de Gauss (32) devient

$$\frac{d^2}{dt^2}\log|f'| = \frac{1-\alpha}{\alpha^2}\frac{f'^2}{f^2}, \tag{42}$$

et (36) donne, puisque $c = \dfrac{1}{\alpha^2}$, $(\log \rho)'' = \dfrac{\rho^2}{\alpha^2}$, ou, en remplaçant ρ par sa valeur

$$\frac{d^2}{dt^2}\log\left|\frac{f'}{f}\right| = \frac{1}{\alpha^2}\frac{f'^2}{f^2}. \tag{43}$$

Retranchant (42) de (43), on obtient $\dfrac{d^2}{dt^2}\log|f| = -\dfrac{1}{\alpha}\dfrac{f'^2}{f^2}$, d'où, par un choix convenable de l'origine des t, $\dfrac{f'}{f} = \dfrac{\alpha}{t}$, ce qui donne

$$f = \left(\frac{t}{h}\right)^{\alpha} \text{ et } \rho^2 = \frac{\alpha^2}{t^2}.$$

Reste à déterminer la forme de $U(u)$ et $V(v)$, par (30), qui, étant donné que $g' = \tau^{\alpha}$ et $\rho^2 = \dfrac{\alpha^2}{t^2}$, s'écrit

$$\frac{U'V'}{(U+V)^2} = \frac{1}{t^2}. \tag{44}$$

Cette équation sera étudiée plus loin (n° **24**); contentons-nous d'indiquer ici le résultat. La solution $U = u$, $V = v$ n'étant pas acceptable, puisque $U+V$ doit être indépendant de $u+v$, on doit prendre

$$U = \frac{1}{ku}, \quad V = \frac{1}{kv}$$

d'où

$$\psi = \frac{1}{k^{\alpha+1}(\alpha+1)}\left(\frac{1}{u}+\frac{1}{v}\right)^{\alpha+1}, \quad f = \left(\frac{t}{h}\right)^{\alpha}.$$

On peut remplacer u et v par hu et hv, à condition de remplacer ψ par $\dfrac{\psi}{h}$, et l'on obtient ainsi

$$\psi = \frac{m}{\alpha+1}\left(\frac{1}{u}+\frac{1}{v}\right)^{\alpha+1}, \quad f = t^{\alpha}, \text{ avec } m = \frac{1}{(\alpha+1)k^{\alpha+1}h^{\alpha}}.$$

Les formes quadratiques de la surface sont

$$ds^2 = 2m\frac{du dv}{(uv)^{\alpha+1}}, \quad \phi = -m\left(\frac{1}{u}+\frac{1}{v}\right)^{\alpha}\left(\frac{du}{u}-\frac{dv}{v}\right)^2. \tag{45}$$

Pour la réalité il faut $m > 0$. Introduisons des paramètres réels r et θ par

$$u = re^{i\theta}, \quad v = re^{-i\theta} \qquad (r > 0)$$

les formes deviennent

$$(45') \qquad ds^2 = 2m \frac{dr^2 + r^2 d\theta^2}{r^{2\alpha+2}}, \quad \phi = 4m \frac{(2\cos\theta)^\alpha}{r^\alpha} d\theta^2.$$

En posant

$$Z_0 = \frac{\sqrt{2m}}{|\alpha|} \frac{1}{u^\alpha}, \quad Z = \frac{\sqrt{2m}}{|\alpha|} \frac{1}{v^\alpha},$$

on applique la surface sur un plan, car le ds^2 devient $dZdZ_0$, ou, en coordonnées polaires $Z = Re^{i\Omega}$, $ds^2 = dR^2 + R^2 d\Omega^2$, avec les relations

$$R = \frac{\sqrt{2m}}{|\alpha| r^\alpha}, \quad \Omega = -\alpha\theta,$$

et la seconde forme s'écrit $\phi = \dfrac{2\sqrt{2m}}{|\alpha|} R \left(2\cos\dfrac{\Omega}{\alpha}\right)^\alpha d\Omega^2$. Les génératrices $\Omega = C$ de la surface sont représentées sur le plan par les demi-droites issues de l'origine, ce qui montre bien qu'il s'agit de cônes. Les droites $\Omega = \pm \pi \dfrac{\alpha}{2}$ du plan représentent des génératrices d'inflexion, si $\alpha > 0$, des génératrices de rebroussement, si $\alpha < 0$. Nous n'envisageons qu'une portion du cône comprise entre deux génératrices de cette sorte, portion qui, sur le plan (v), serait représentée sur le demi-plan de droite; le cône entier s'obtient à partir d'une telle portion par des symétries relativement à des plans ou à des droites.

Si l'on coupe le cône par la sphère unité centrée à son sommet, on trouve une courbe dont la courbure géodésique, relativement à la sphère, est identique à la courbure normale de la même courbe relativement au cône, c'est-à-dire à $\dfrac{2\sqrt{2m}}{|\alpha|}\left(2\cos\dfrac{\Omega}{\alpha}\right)^\alpha$. Comme l'arc élémentaire de cette courbe est $ds = d\Omega$, son équation intrinsèque est

$$(46) \qquad \frac{1}{\rho_g} = \frac{2\sqrt{2m}}{|\alpha|}\left(2\cos\frac{s}{\alpha}\right)^\alpha.$$

On constate bien qu' elle présente, pour $s = \pm \dfrac{\pi\alpha}{2}$, des inflexions si $\alpha > 0$, des rebroussements si $\alpha < 0$.

Nous savons que, sur ce cône, les lignes $H = C$, $\psi = C$ sont des courbes isothermes. Dans l'application, elles deviennent des courbes isothermes du plan, qu'il est facile de déterminer. Les courbes $H = C$ ont pour équation $u + v = C$, ou $r\cos\theta = C$, ce qui devient, en coordonnées R et Ω, $R^{-\frac{1}{\alpha}} \cos\dfrac{\Omega}{\alpha} = C$; ce sont les courbes obtenues en égalant à une constante la partie réelle de la fonction analytique $Z^{-\frac{1}{\alpha}}$. L'équation des courbes $\psi = C$ est $\dfrac{1}{u} + \dfrac{1}{v} = C$, ou

$\frac{1}{r}\cos\theta = 0$, c'est-à-dire $R^{\frac{1}{a}}\cos\frac{\Omega}{a} = C$. Sur le plan, les courbes $\psi = C$ sont les inverses des courbes $H = C$, dans une inversion de pôle 0. Cette inversion conserve les droites passant par 0 et les cercles centrés en 0, images des lignes de courbure: on vérifie que les lignes de courbure bissectent les lignes $H = C$ et $\psi = C$.

En rapprochant l'équation des courbes $\psi = C$ de l'expression de $\frac{1}{\rho_g}$ pour la courbe sphérique directrice du cône, on voit que, sur ces courbes, R varie proportionnellement à ρ_g, d'où cette proposition: *si l'on porte, à partir du sommet, sur chaque génératrice du cône, une longueur égale au rayon de courbure géodésique de la courbe sphérique intersection du cône et de la sphère unité, on obtient une courbe primitive du cône; les autres courbes primitives sont homothétiques à celle-là, les courbes d'égale courbure moyenne sont les inverses des précédentes, le centre d'homothétie et le pôle d'inversion étant le sommet du cône.*

On peut d'ailleurs remarquer que, sur tout cône, si l'on porte à partir du sommet, sur chaque génératrice, une longueur inverse du rayon de courbure géodésique de la courbe sphérique déterminée par le cône sur la sphère unité, la courbe obtenue est une ligne d'égale courbure moyenne. Il y a là un moyen de déterminer directement les cônes à courbure moyenne isotherme, et, par conséquent, de contrôler nos calculs. La fonction $f(s)$, qui figure dans l'équation intrinsèque $\rho_g = f(s)$ de la courbe sphérique directrice du cône, doit être telle que, dans le plan $Z = Re^{i\Omega}$, les courbes $Rf(\Omega) = C$, soient isothermes; on retrouve bien, comme le montre un calcul facile, $\frac{1}{\rho_g} = k\left(\cos\frac{s}{a}\right)^a$.

IV. Surfaces W applicables sur des surfaces de révolution

17. Le ds^2 d'une surface applicable sur une surface de révolution peut s'écrire $ds^2 = 2\, F(u+v)dudv$. La courbure totale $\mathbf{K}$ sera donc, elle aussi, fonction de $u + v$; nous supposerons que $\mathbf{K}$ n'est pas une constante.

Supposons en outre que la surface soit W; la courbure moyenne sera fonction de $\mathbf{K}$, donc de $u + v$, ce qui revient à dire que la surface sera à courbure moyenne isotherme, ou constante. N'examinons que le cas où la courbure moyenne est variable, $H = f(u + v)$. La fonction primitive de la surface sera donc telle que $\psi_{uv} = Ff'$; donc ψ_{uv} ne dépendra que de $u + v$. Par ailleurs, l'asphéricité A, dont l'expression est $\frac{\sqrt{\psi_u\psi_v}}{\psi_{uv}} f'$, dépend, elle aussi, uniquement de $u + v$; donc, le produit $\psi_u\psi_v$ est fonction de $u + v$. D'où ce premier résultat: *si une surface W, à courbure totale et à courbure moyenne variables, est applicable sur une surface de révolution, c'est une surface HI, sur laquelle ψ_{uv} et $\psi_u\psi_v$ sont fonction de $u + v$, comme H.*

Il est facile de trouver toutes les surfaces *HI* dont la fonction primitive jouit de ces deux propriétés; il suffit d'utiliser l'identité suivante, que nous avons déjà signalée ailleurs:

$$(47)\qquad \psi_u\left(\frac{\psi_{uv}}{\psi_u\psi_v}\right)_v + \psi_v\left(\frac{\psi_{uv}}{\psi_u\psi_v}\right)_u = (\log \psi_u\psi_v)_{uv} - 2\frac{\psi_{uv}^2}{\psi_u\psi_v}.$$

18. Supposons que les deux fonctions de $u + v$, ψ_{uv} et $\psi_u\psi_v$ soient linéairement indépendantes, autrement dit, que le rapport $\frac{\psi_{uv}}{\psi_u\psi_v}$ ne soit pas constant. La formule (47) fournit alors une relation linéaire entre ψ_u et ψ_v, dont les coefficients ne dépendent que de $u + v$. Cette relation, jointe à l'expression de $\psi_u\psi_v$ en $u + v$, montre que ψ_u et ψ_v sont séparément fonction de $u + v$. Comme ψ_{uv} ne dépend, lui non plus, que de $u + v$, on a nécessairement des expressions telles que

$$\psi_u = g(u + v) - ia, \quad \psi_v = g(u + v) + ia.$$

On montre que ces surfaces sont, en général, des *hélicoïdes*, ou, si $a = 0$, des *surfaces de révolution*. Le cas particulier où l'on aurait $\psi_{uv} = Cf'$ correspondrait à un *cylindre* (n° **12**), mais il ne doit pas être retenu, puisque nous supposons la surface à courbure totale variable.

19. Supposons au contraire que le rapport $\frac{\psi_{uv}}{\psi_u\psi_v}$ soit constant; c'est une condition déjà rencontrée (n° **13** A), qui caractérise, comme nous le verrons, les surfaces d'Ossian Bonnet. Nous obtenons donc le théorème: *si une surface W, à courbures totale et moyenne variables, est applicable sur une surface de révolution, c'est, ou bien un hélicoïde, ou bien une surface de révolution, ou bien une surface d'Ossian Bonnet.*

V. Surfaces d'Ossian Bonnet

20. Les surfaces d'Ossian Bonnet sont les surfaces susceptibles d'une infinité de déformations avec conservation des courbures principales. Nous établissons d'abord leurs équations différentielles.

Soient S et $\overline{S}$ deux surfaces applicables l'une sur l'autre avec conservation des courbures principales. Leurs éléments linéaires respectifs sont $\frac{2\omega_1\omega_2}{A}$ et $\frac{2\overline{\omega}_1\overline{\omega}_2}{\overline{A}}$. Comme $\overline{A} = A$, par hypothèse, l'isométrie exige

$$(48)\qquad \overline{\omega}_1\overline{\omega}_2 = \omega_1\omega_2.$$

La formule de Codazzi (13) donne d'autre part, puisque $\overline{H} = H$

$$(49)\qquad \overline{r\omega_1} + \overline{s\omega_2} = r\omega_1 + s\omega_2.$$

Une surface est surface d'O. Bonnet si ces deux équations en $\overline{\omega}_1$, $\overline{\omega}_2$ ont une

infinité de solutions. L'équation (48) exige, compte tenu de la réalité, que

(50) $$\bar{\omega}_1 = e^{i\theta}\omega_1, \qquad \bar{\omega}_2 = e^{-i\theta}\omega_2,$$

où θ est l'angle que fait, après application, la première ligne de courbure de $\bar{S}$ avec la première ligne de courbure de S. (49) donne ensuite

(51) $$\bar{r} = re^{-i\theta}, \quad \bar{s} = se^{i\theta}.$$

Mais en différentiant extérieurement (50), et remarquant que $[\bar{\omega}_1\bar{\omega}_2] = [\omega_1\omega_2]$, on obtient

$$\bar{r} = (r - i\theta_2)e^{i\theta}, \quad \bar{s} = (s + i\theta_1)e^{-i\theta}$$

d'où, en éliminant $\bar{r}$ et $\bar{s}$ à l'aide de (51),

$$i\theta_1 = s(e^{2i\theta} - 1) \quad i\theta_2 = r(1 - e^{-2i\theta})$$

et enfin

(52) $$id\theta = s(e^{2i\theta} - 1)\omega_1 + r(1 - e^{-2i\theta})\,\omega_2.$$

Il ne reste plus qu'à exprimer que cette équation de Pfaff est complètement intégrable, ce qui donne

(53) $$r_1 + rs = 0, \quad s_2 + rs = 0;$$

telles sont les deux équations, du 4e ordre, des surfaces d'O. Bonnet.

21. Le théorème d'O. Bonnet, d'après lequel ses surfaces sont isothermiques, se lit sur les formules (53), car on en tire $r_1 = s_2$, ce qui est l'équation des surfaces isothermiques (n° 8). On en tire aussi $s = -\dfrac{r_1}{r}$, $r = -\dfrac{s_2}{s}$.

Notre forme χ (n° 4) s'écrit donc

$$\chi = \left(-\frac{r_1}{r} - \frac{s_1}{s}\right)\omega_1 + \left(-\frac{s_2}{s} - \frac{r_2}{r}\right)\omega_2 = -d\log rs.$$

C'est une différentielle exacte, donc (n° 4), les surfaces en question sont des surfaces *HI*. En se reportant aux formules (21) on voit que

$$\chi = d\log\frac{\sqrt{\psi_u\psi_v}}{\psi_{uv}} = -d\log rs \quad \text{et} \quad \frac{1}{rs} = 4\frac{(\psi_u\psi_v)^{\frac{3}{2}}}{\psi_{uv}^2}$$

d'où, en éliminant rs,

(54) $$\frac{\psi_{uv}}{\psi_u\psi_v} = m.$$

Réciproquement, toute surface *HI* dont la fonction primitive a ses deux paramètres différentiels proportionnels est une surface d'O. Bonnet. En effet, de (54) on déduit $r = -\dfrac{m}{2}\sqrt{\psi_v}$, $s = -\dfrac{m}{2}\sqrt{\psi_u}$ d'où

$$r_1 = -\frac{m^2}{4}\sqrt{\psi_u\psi_v} = s_2 = -rs.$$

En définitive, *la condition nécessaire et suffisante pour qu'une surface soit surface d'O. Bonnet, c'est qu'elle soit une surface HI et que sa fonction primitive ait ses deux paramètres différentiels proportionnels.*

On doit remarquer que les équations (53) sont satisfaites si $r = 0, s = 0$; (52) donne alors $\theta =$ const. Les surfaces d'O. Bonnet correspondantes sont *à courbure moyenne constante*: ce sont les surfaces d'O. Bonnet *de première classe*; nous ne nous en occuperons pas.

22. Revenons sur la propriété que possède toute surface d'O. B. d'être isothermique. De (54), on tire par intégration

$$\psi = -\frac{1}{m}\log(U+V), \tag{55}$$

où U et V sont, pour la réalité, imaginaires conjuguées. La forme isothermique $s\omega_1 + r\omega_2$ s'écrit donc $-\frac{1}{2}\frac{\psi_{uv}}{\psi_u\psi_v}d\psi = -\frac{m}{2}d\psi$, et le facteur intégrant commun à ω_1 et ω_2, qui est en général $e^{\int s\omega_1 + r\omega_2}$, devient $e^{-\frac{m}{2}\psi} = \sqrt{U+V}$. Nous poserons $U = P - iQ$, $V = P + iQ$, et nous nous restreindrons à une région de la surface où $P > 0$. $P = \dfrac{U+V}{2}$ est une fonction harmonique sur la surface, et l'on a

$$s\omega_1 + r\omega_2 = \frac{1}{2}\frac{dP}{P}. \tag{56}$$

Réciproquement, s'il existe une fonction harmonique P telle que la forme isothermique puisse s'écrire ainsi, on aura

$$P_1 = 2sP,\ P_2 = 2rP,\ P_{12} = 2s_2P + 4rsP,$$

et enfin

$$P_{12} - rP_1 = 2(s_2 + rs)P.$$

Or, puisque P est harmonique, $P_{12} - rP_1 = 0$; donc, sur de telles surfaces, $s_2 + rs = 0$; on montrerait de même, en formant $P_{21} - sP_2$, que $r_1 + rs = 0$, donc *la condition nécessaire et suffisante pour qu'une surface soit surface d'O. Bonnet, c'est que la forme isothermique soit la demi-différentielle logarithmique d'une fonction harmonique.*

23. Les surfaces à courbure moyenne constante sont, nous l'avons dit, les surfaces d'O.B. de première classe. Les autres surfaces d'O.B, à courbure moyenne variable, se répartissent en deux classes, comme le montre l'étude de l'équation de Gauss. Puisqu'ici, $\psi = -\frac{1}{m}\log(U+V)$, on a

$$\psi_{uv} = \frac{U'V'}{m(U+V)^2} \quad \text{et} \quad \frac{\partial^2}{\partial u\partial v}\log|\psi_{uv}| = -2\psi_{uv}.$$

L'équation (23) devient alors

$$\frac{d^2}{dt^2}\log|f'| = \psi_{uv}\left(\frac{f^2}{f'} + 2m\right) - \frac{f'}{m}. \tag{57}$$

Si on lui applique l'opération $D = \frac{\partial}{\partial u} - \frac{\partial}{\partial v}$, on obtient

$$0 = D\psi_{uv} \cdot \left(\frac{f^2}{f'} + 2m\right) \tag{58}$$

d'où deux possibilités:

ou bien $\frac{f^2}{f'} + 2m = 0$, qui donne $f = \frac{2m}{t + C}$. Cette fonction satisfait à (57) sans qu'aucune condition soit imposée aux fonctions U et V: ce sont les surfaces d'O.B. *de deuxième classe.* On peut déterminer complètement l'expression de leurs coordonnées en fonction de u, v et de deux fonctions arbitraires, mais elles sont imaginaires, comme l'a montré M. E. Cartan;

ou bien $D\psi_{uv} = 0$: ψ_{uv} n'est fonction que de $u + v$: ce sont les surfaces d'O.B. *de troisième classe*, que nous allons étudier.

24. Les surfaces d'O. Bonnet de troisième classe jouissent de plusieurs propriétés qui sont évidentes sur nos formules. Leur élément linéaire étant $2\frac{\psi_{uv}}{f'}\,dudv$, comme pour toute surface HI, et ψ_{uv} ne dépendant que de $u + v$, comme f', *elles sont applicables sur des surfaces de révolution* (n° **17**). En outre, en se reportant à l'expression (21) de A, et en tenant compte que $\psi_u\psi_v = \frac{\psi_{uv}}{m}$, on voit que A n'est fonction que de $u + v$, comme H, si bien que A est fonction de H et que, par conséquent, *ces surfaces sont des surfaces W.* Ces propriétés sont du reste bien connues.

D'après la formule (55), les courbes primitives $\psi = C$ sont des courbes isothermes. Conjointement avec leurs trajectoires orthogonales, elles forment un réseau auquel est attachée la variable complexe V, en ce sens qu'on obtient les courbes $\psi = C$ et leurs trajectoires orthogonales en égalant à une constante soit la partie réelle, soit la partie imaginaire de V. Nous poserons $V = P + iQ$; cette variable, qui est attachée aux courbes primitives dans le même sens que la variable $v = p + iq$ est attachée aux lignes d'égale courbure moyenne, sera appelée la *variable primitive.* Puisque V et v sont deux intégrales premières de $\omega_2 = 0$, il y a une relation analytique $V(v)$, que nous appellerons la *relation primitive.*

Sur ces surfaces, on a, d'après (55).

$$\psi_{uv} = \frac{U'V'}{m(U + V)^2}; \tag{59}$$

ce doit être une fonction de $u + v$, donc les dérivées logarithmiques par rapport à u et par rapport à v sont égales, ce qui donne

$$\frac{U''}{U'} - \frac{2U'}{U + V} = \frac{V''}{V'} - \frac{2V'}{U + V} \tag{60}$$

ou, en posant, pour abaisser l'ordre, $U' = \lambda(U)$, $V' = \mu(V)$,

$$(\lambda' - \mu')(U + V) = 2(\lambda - \mu). \tag{61}$$

Des dérivations par rapport à U, puis par rapport à V, donnent

(62) $$\lambda''(U + V) = \mu''(U + V) = \lambda' + \mu'.$$

On en déduit

$$\lambda'' = \mu'' = 2a \qquad (a, \text{ constante réelle})$$

d'où

$$\lambda' = 2aU + b_1, \ \mu' = 2aV + b_2 \qquad (b_1, b_2 \text{ imaginaires conjuguées}).$$

Portant dans (62), on obtient $b_1 + b_2 = 0$, donc

$$b_1 = 2ib, \quad b_2 = -2ib \qquad (b, \text{ constante réelle})$$

et (61) donne ensuite

$$\lambda - aU^2 - 2ibU = \mu - aV^2 + 2ibV,$$

d'où

$$\lambda = aU^2 + 2ibU + c, \ \mu = aV^2 - 2ibV + c \qquad (c, \text{ constante réelle}).$$

La relation différentielle primitive est donc

(63) $$\frac{dV}{dv} = aV^2 - 2ibV + c.$$

25. Nous distinguerons trois types de surfaces, suivant la nature des racines V_1, V_2 du trinôme $aV^2 - 2ibV + c$, où a, b et c sont réels:

type **A** : $b^2 + ac > 0$, 2 racines distinctes, imaginaires pures;

type **B** : $b^2 + ac < 0$, 2 racines distinctes, symétriques par rapport à l'axe imaginaire;

type **C** : $b^2 + ac = 0$, 2 racines confondues sur l'axe imaginaire.

Les types **A** et **C** ont deux réalisations: la *normale*, $a \neq 0$, et la *spéciale*, $a = 0$. Dans le type **A** spécial, une des racines est rejetée à l'infini; dans le type **C** spécial, les deux racines sont infinies.

On peut simplifier la relation primitive différentielle (63) en utilisant le fait que les variables complexes V et v ne sont pas parfaitement définies. Pour V, c'est seulement par sa partie réelle qu'elle intervient, puisque, seule, figure dans les formules la somme $U + V = 2P$; on peut ajouter à V une constante imaginaire pure, à condition de retrancher de U la même quantité. De plus, si l'on multiplie U et V par une même constante positive, ψ est simplement augmentée d'une quantité constante, ce qui ne change rien à la surface. Quant à v, on peut la remplacer par $av + b + ic$ (a, b, c réels), à condition de remplacer u par $au + b - ic$, ce qui remplacera $u + v$ par $a(u + v) + 2b$. Il ne faut pas oublier que, dans ce changement, ψ ne reste pas invariante (n° 6), mais devient $\overline{\psi} = \dfrac{\psi}{a}$; en particulier, si u et v sont remplacées par $-u$ et $-v$, ψ devient $-\psi$. On obtient de la sorte les formes réduites:

A normal, $$dv = \frac{dV}{1 + V^2}, \qquad V = \operatorname{tg} v$$

A spécial, $$dv = \frac{dV}{2iV}, \qquad V = -ie^{2iv}$$

$$\mathbf{B}, \qquad dv = \frac{dV}{1 - V^2}, \qquad V = \operatorname{th} v$$

$$\mathbf{C} \text{ normal}, \qquad dv = -\frac{dV}{V^2}, \qquad V = \frac{1}{v}$$

$$\mathbf{C} \text{ spécial}, \qquad dv = dV, \qquad V = v.$$

On a intégré de façon que les axes imaginaires se correspondent, ainsi que les demi-plans positifs, dans les plans complexes (v) et (V).

Les expressions correspondantes de ψ_{uv} sont

$$\text{(64)} \qquad \mathbf{A},\ \frac{1}{m \sin^2 t}; \quad \mathbf{B},\ \frac{1}{m \operatorname{sh}^2 t}; \quad \mathbf{C},\ \frac{1}{mt^2};$$

et l'équation de Gauss (57), que doit vérifier $H = f(t)$, revêt aussi trois formes différentes, suivant le type considéré.

26. La relation (63) montre que les lignes du plan (V) le long desquelles dv est réel sont des cercles (ou droites) orthogonaux à l'axe imaginaire. Donc, quel que soit le type considéré, *si l'on fait la carte de la surface d'O.B. sur le plan (V), les trajectoires orthogonales des lignes $H = C$ ont pour image un faisceau de cercles (ou droites) orthogonaux à l'axe imaginaire; les courbes $H = C$ elles-mêmes sont représentées par le faisceau orthogonal au précédent.*

Le faisceau de cercles qui représente les trajectoires orthogonales des lignes d'égale courbure moyenne peut avoir ses points de Poncelet sur l'axe imaginaire, distincts (type **A**), ou bien ses points de base symétriques par rapport à l'axe imaginaire (type **B**), ou bien ses points de base confondus sur l'axe imaginaire, l'axe radical étant perpendiculaire à l'axe imaginaire (type **C**). Dans le type **A** normal, les courbes $H = C$ sont représentées par des arcs de cercle limités aux points V_1, V_2; dans le type **A** spécial, elles sont représentées par des droites rayonnant de V_1, V_2 étant à l'infini. Dans le type **B**, les courbes $H = C$ ont pour image un faisceau de cercles ayant comme points de Poncelet V_1 et V_2, symétriques par rapport à l'axe imaginaire. Enfin, dans le type **C**, les courbes $H = C$ sont des cercles tangents en $V_1 = V_2$ à l'axe imaginaire, ou, si le type est spécial, des droites parallèles à l'axe imaginaire.

27. Cherchons maintenant la carte des lignes de courbure sur le plan (V). Nous savons (6) qu'elles bissectent les courbes primitives et les courbes $H = C$. Or, sur le plan (V), les lignes $\psi = C$ sont les parallèles à l'axe imaginaire, et les lignes $H = C$ forment un faisceau de cercles comme on vient de le voir. Des considérations élémentaires montrent qu'en conséquence *les lignes de courbure seront, sur la carte, des coniques homofocales*, les foyers étant les points V_1 et V_2. Ce seront des coniques à centre dans le type **A** normal et dans le type **B**, des paraboles homofocales, ayant pour axe l'axe imaginaire, dans le type **A** spécial; dans le type **C** normal, ce seront les demi-droites issues du point $V_1 = V_2$ de l'axe imaginaire, et les cercles centrés en ce point; dans le type **C** spécial, ce seront des parallèles aux axes.

Nous avons montré (n° **8**) comment on attache une variable complexe z au réseau des lignes de courbure d'une surface isothermique. Ici, comme

$$e^{\int s\omega_1 + r\omega_2} = \sqrt{U + V}, \text{ et que } \omega_2 = \sqrt{\psi_v}dv = \sqrt{-\frac{1}{m}\frac{V'}{U + V}}\, dv,$$

on pourra prendre

$$dz = \sqrt{V'}\, dv \text{ ou } dz = i\sqrt{V'}\, dv, \tag{65}$$

suivant que m est négatif ou positif; en toute hypothèse, on a $dz^2 = \pm\, dVdv$, Si, dans (65), on remplace dv au moyen de (63) on obtient

$$dz = \frac{dV}{\sqrt{aV^2 - 2ibV + c}} \text{ (si } m < 0), \quad dz = \frac{idV}{\sqrt{aV^2 - 2ibV + c}} \text{ (si } m > 0), \tag{66}$$

ce qui confirme que les lignes de courbure, $y = C$, $x = C$, ont pour carte, dans le plan (V), des coniques homofocales. Les formules (66) donnent en outre, dans chaque cas, par intégration, l'expression de V en z, c'est-à-dire, de P et Q en x, y.

28. La forme de la fonction $P(x, y)$ est remarquable: *P est le produit d'une fonction de x par une fonction de y.* Cette propriété, dont M. E. Cartan a tiré un grand parti, découle d'une proposition plus générale établie antérieurement (n° **11**). Ici, on a (n° **22**),

$$s\omega_1 + r\omega_2 = \frac{1}{2}\frac{dP}{P},$$

donc, en vertu de la propriété rappelée, $P = X(x)Y(y)$. Comme P est une fonction harmonique, les formes possibles de $X(x)$ et $Y(y)$ sont très limitées.

Nous ne développerons pas davantage la théorie des surfaces d'Ossian Bonnet qui, étant donné son grand intérêt, mérite une étude à part. Notre intention était seulement de les présenter ici à titre de spécimen remarquable des surfaces à courbure moyenne isotherme.

Issy-les-Moulineaux (*Seine*)

DISCRETE SPACE-TIME AND INTEGRAL LORENTZ TRANSFORMATIONS

ALFRED SCHILD

Introduction. Modern physical theory, both classical and quantal, faces serious difficulties which arise from the divergence of certain integrals. Perhaps the best known of these "infinities" is the self-energy of the point electron. Most of the simpler devices used to eliminate the infinities, such as the introduction of a finite electron radius, are non-relativistic and must therefore be rejected. Relativistic theories[1] which do avoid some or all of the infinities are very complicated and often suffer from difficulty in physical interpretation.

The idea of introducing discreteness into space and time has occasionally been considered.[2] It seems likely that a physical theory based on a discrete space-time background will be free of the infinities which trouble contemporary quantum mechanics. The objection which is usually raised against such discrete schemes is that they are not invariant under the Lorentz group. The purpose of this investigation is to show that *there is a simple model of discrete space-time which, although not invariant under all Lorentz transformations, does admit a surprisingly large number of Lorentz transformations.* This group of transformations is, in fact, sufficiently large to make doubtful the validity of most physical objections raised against discrete space-times.

Apart from the physical speculations in the introduction, this paper is of a purely mathematical nature. We consider all events in Minkowski space-time whose four coordinates t, x, y, z are integers. (The velocity of light is taken as unity.) These events form a "cubic lattice"[3] in space-time. We first investigate the null lines which join lattice points, then the Lorentz transformations which leave the cubic lattice as a whole invariant. We shall call these *integral null lines* and *integral Lorentz transformations*, respectively. We also consider the time-like lines through lattice points which are mapped into lines parallel to the t-axis by an integral Lorentz transformation. These lines will be called *integral time lines*.

It may be noted that our model of discrete space-time involves a *fundamental length*[4] ϵ, namely, the least non-zero interval between lattice points. In the present investigation this fundamental distance has been chosen as the unit of length. In any physical theory based on our model, ϵ would probably be of the general order of magnitude of the classical electron radius (approximately 10^{-13} cm.).

Received January 26, 1948.

[1] G. Wentzel, *Rev. Mod. Phys.*, vol. 19 (1947), 1-18.

[2] V. Ambarzumian and D. Iwanenko, *Z. f. Phys.*, vol. 64 (1930), 563-567; L. Silberstein, "Discrete Space-Time," University of Toronto Studies, Physics Series (1936). For a short outline of the present paper, see *Phys. Rev.*, vol. 73 (1948), 414-415.

[3] "Hypercubic" would be the appropriate adjective—but we shall retain the shorter form.

[4] Cf. W. Heisenberg, *Ann. Phys.*, vol. 32 (1938), 20-33.

There are two attractive possibilities for making a first rough attempt at introducing physical theory on our discrete space-time background. The motion of a particle may be assumed to consist of a temporally ordered sequence of lattice points such that successive lattice points are joined by (*a*) integral null lines, or (*b*) integral time lines. In case (*a*), a particle always moves with an instantaneous velocity equal to the velocity of light, but it changes direction rapidly so that its average velocity can be quite low. This zigzag motion has a striking resemblance to some of the features of the Dirac electron.[5] Case (*b*) is rather similar to (*a*). The main difference is that the instantaneous velocity of a particle may now be zero; however it is interesting to note that the non-zero velocities associated with integral time lines are all very high and exceed 0.86 times the velocity of light (Sec. **8**).

Two of the results which we obtain are particularly striking. The first states that the spatial projections of integral null lines are dense (Sec. **4**). This means that particles, whose motion is of the type (*a*) above, can have instantaneous velocities in practically any direction of space. We shall also show that all integral null lines are equivalent in the sense that, given any two integral null lines, an integral Lorentz transformation can be found which maps one into the other (Sec. **7**).

The second result states that spatial projections of integral time lines are dense (Sec. **8**). This means that particles, whose motion is of the type (*b*) above, can have instantaneous velocities in practically any direction of space.

It is obvious that the cubic lattice which we are considering is invariant under all *translations* which map one lattice point into another. In this sense our discrete model of space-time is *homogeneous*. The two results stated above show that our model possesses also a large measure of *spatial isotropy*.

Of any physical theory based on our model of discrete space-time we require invariance under integral Lorentz transformations. The integral Lorentz transformations are independent of the fundamental length ϵ. Thus in the limit when ϵ tends to zero we expect the resulting equations of the physical theory to remain invariant under integral Lorentz transformations, although the background is now continuous Minkowski space-time. If the limiting equations are at all simple they are almost certain to be invariant under *all* Lorentz transformations, since it is difficult to visualize equations in continuous space-time which are invariant under as substantial a subgroup of Lorentz transformations as that considered here without these equations being completely Lorentz invariant. Thus it is reasonable to hope that equations based on our discrete space-time model might be found which, in the limit $\epsilon \to 0$, take the form of the equations of "continuous" relativistic physics, e.g. Maxwell's equations, Lorentz's equations of motion, and Dirac's equations for the electron. These equations of "continuous" physics would be a valid approximation for macroscopic phenomena and even for atomic and molecular

[5] E. Schrödinger, *Sitz. Ber. Preuss. Akad. Wiss.*, vol. 24 (1930), 418-428.

theory—but they would not be appropriate for the description of nuclear phenomena or the theory of elementary particles.

It is clear that we have merely chosen the simplest discrete model of space-time. Other regular point lattices in space-time might be considered and perhaps found more useful. In most essentials, however, these lattices would behave much the same as the cubic lattice studied here. For example, the Lorentz transformations which leave any such lattice invariant would all be associated with high velocities.

1. **Gaussian Integers.** In this section we collect some well-known definitions and theorems concerning Gaussian integers which will be used in the sequel.

A Gaussian integer is a complex number $a + ib$ whose real part a and imaginary part b are both integers. A real Gaussian integer is an ordinary integer. Gaussian integers can be added, subtracted and multiplied to yield other Gaussian integers; they form an *integral domain*. Here and in the following we shall refer to Gaussian integers simply as "integers." Sometimes, when we are dealing with ordinary integers, we shall add the adjective "real," but usually it will be clear from the context whether integers are real or complex (Gaussian).

The complex conjugate of $c = a + ib$ will be denoted by $\bar{c} = a - ib$; the absolute value of c by $|c| = +(a^2 + b^2)^{\frac{1}{2}}$.

A *unit* is an integer which divides all integers. There are exactly four units in the Gaussian integral domain: ± 1, and $\pm i$.

A *prime* p is an integer which is divisible only by the four units ± 1, $\pm i$, and by $\pm p$, $\pm ip$. Two integers are *relatively prime* if their only common factors are units. Similarly, a set of integers with units as their only common factors will be called *primitive*; thus a primitive vector is a vector whose components form a primitive set of integers.

One of the most important properties of the Gaussian integral domain is that it admits of *unique factorization* into primes.[6] By this is meant the following: An integer a can be written in the form

$$a = p_1 p_2 \ldots p_r, \tag{1.01}$$

where the p_i are primes other than units; if it can also be written in the form

$$a = q_1 q_2 \ldots q_s, \tag{1.02}$$

where the q_i are primes other than units, then $r = s$, and, for a suitable relabelling of the factors $q_1, \ldots, q_r$ we have

$$p_1 = u_1 q_1,\ p_2 = u_2 q_2, \ldots, p_r = u_r q_r, \tag{1.03}$$

where $u_1, u_2, \ldots, u_r$ are units.

We shall apply the term *real prime* to a prime, as defined above, which is real. This definition does not agree with the usual one for real integers in

[6] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (Oxford, 1938), 184, Theorem 215.

which the criterion is the absence of *real* non-trivial factors; thus $2 = (1 + i)(1 - i)$ and $5 = (2 + i)(2-i)$ are not real primes as we have defined the term. It is clear that any real integer p can be written in the form

$$p = q\, a\, \bar{a}, \tag{1.04}$$

where a, $\bar{a}$ are complex conjugate integers, and where q is the product of those real primes, each taken once, which are factors of p an odd number of times. This q is the real integer of least magnitude for which a decomposition of p in the form (1.04) is possible; apart from sign, q is uniquely determined by p. Since 2 is not a real prime, q must be odd.

Although we shall not require it in the sequel, we add the well-known theorem[7] that among the numbers 2, 3, 5, 7, . . . (which are usually called primes) those and only those of the form $4n + 3$, n being a real integer, are real primes.

An integer $a + ib$, where a and b are real, will be called *even* if a and b are either both even or both odd in the conventional sense; $a + ib$ will be called *odd* if one of a, b is odd and the other even. It is immediately obvious that for real integers our definitions of the terms even and odd coincide with the conventional meaning. The following facts are easily proved:

An even integer is divisible by the prime $1 + i$, an odd integer is not. (Note that the primes $1 - i$, $-1 + i$, $-1 - i$ differ from $1 + i$ only by the unit factors $-i$, i, -1, respectively.) The sum of two integers is even if the integers are both even or both odd; otherwise the sum is odd. The product of two integers is odd only if both factors are odd; otherwise it is even. These rules are easy to remember as they are all familiar from the conventional properties of even and odd real integers; in the case of Gaussian integers the conventional role of 2 is taken over by the prime $1 + i$ which is a repeated factor of 2:

$$2 = -i(1 + i)^2. \tag{1.05}$$

We require some further theorems of which the first is a standard result:

If a and b are relatively prime, then there exist integers l and m such that

$$la - mb = 1. \tag{1.06}$$

Conversely (1.06) implies that a and b are relatively prime.

Equation (1.06) may be regarded as a diophantine equation for the unknown integers l and m. If l, m is a particular solution, then the general solution is $l + pb$, $m + pa$, where p is an arbitrary integer. Thus the general solution of (1.06), if relatively prime integers a and b are assigned, involves one discrete complex parameter p or two discrete real parameters. If (1.06) is now regarded as a diophantine equation for the four unknown integers a, b, l, m, then there is a *discrete sixfold infinity of solutions*, since the complex integers a and b can be chosen arbitrarily except for the restriction that they be relatively prime.

In (1.06), a and b are either both odd or else one of them is odd and the other even. If a and b are both odd, then one of l, m must be odd and the other even, so that $a + b + l + m$ is odd.

[7]Hardy and Wright, *Theory of Numbers*, 219, Theorem 252.

In the other case let us, for the sake of definiteness, take a even and b odd. Then one of two possibilities can arise: (i) l and m are both odd, so that $a + b + l + m$ is odd; (ii) l is even and m is odd, so that $a + b + l + m$ is even. Given a solution of (1.06) in which a is even and b, l, m are odd, then

(1.07) $$(l + b)a - (m + a)b = 1,$$

and $a, (l + b)$ are even, b, $(m + a)$ are odd. We easily deduce the results:

If two relatively prime integers a and b are assigned, a being even and b odd, then there exists an even integer l and an odd integer m, satisfying (1.06).

Equation (1.06) has a discrete sixfold infinity of solutions in integers, such that $a + b + l + m$ is even.

2. Spinors and Tensors. We give here a short survey of the spinor calculus[8] in the form in which it will be applied to our problem.

In a complex plane (i.e. a plane with two complex coordinates), called the *spin space*, vectors and tensors are defined by their usual transformation properties. Thus

(2.01) $$c'^{\alpha} = \lambda_{\beta}{}^{\alpha}\, c^{\beta},$$

where the $\lambda_{\beta}{}^{\alpha}$ are constants, is the transformation equation of a *contravariant spinvector* c^{α}. Greek suffixes range over 1, 2 and the usual range convention and summation convention for repeated suffixes are assumed. The components of c^{α} are complex and we denote their complex conjugates by $c^{\dot\alpha}$. Then, obviously,

(2.02) $$c'^{\dot\alpha} = \bar\lambda_{\beta}{}^{\alpha}\, c^{\dot\beta},$$

where $\lambda_{\beta}{}^{\alpha}$ denotes the complex conjugate of $\lambda_{\beta}{}^{\alpha}$. Expressions such as $a^{\dot\alpha\beta}$, $a^{\alpha\dot\beta}$, $a^{\dot\alpha\beta\gamma}$, etc., are called *contravariant spintensors* or *spinors* if they have, respectively, the same transformation equations as $c^{\dot\alpha}c^{\beta}$, $c^{\alpha}c^{\dot\beta}$, $c^{\dot\alpha}c^{\beta}c^{\gamma}$, etc. If, in a spinor, dots are placed on undotted suffixes and the dots removed from dotted suffixes, then the resulting spinor denotes the complex conjugate of the original spinor; thus

(2.03) $$a^{\alpha\dot\beta} = \overline{a^{\dot\alpha\beta}}, \quad a^{\dot\alpha\beta\gamma} = \overline{a^{\alpha\dot\beta\dot\gamma}}, \text{ etc.}$$

A spintensor $a^{\dot\alpha\beta}$ which has the symmetry property

(2.04) $$a^{\dot\alpha\beta} = a^{\beta\dot\alpha},$$

or, equivalently,

(2.05) $$a^{\dot 1 1} = \overline{a^{\dot 1 1}}, \quad a^{\dot 1 2} = \overline{a^{\dot 2 1}}, \quad a^{\dot 2 2} = \overline{a^{\dot 2 2}},$$

is said to be *Hermitian*.

Let us now consider *Minkowski space-time* which is a flat real 4-space with coordinates

(2.06) $$(t, x, y, z) \equiv (x^0, x^1, x^2, x^3),$$

and with metric tensor

---

[8] O. Laporte and G. E. Uhlenbeck, *Phys. Rev.*, vol. 37 (1931), 1381; L. Infeld, *Phys. Zeitschrift*, vol. 33 (1932), 475. For an early use of a similar technique see also E. Goursat, *Ann. École Norm.* (3), vol. 6 (1889), 20, § 5.

(2.07) $$g_{00} = 1,\ g_{11} = g_{22} = g_{33} = -1,\ g_{rs} = 0 \text{ for } r \neq s.$$

Latin suffixes range over 0, 1, 2, 3 and the range and summation conventions are assumed. The ***Lorentz transformations*** are the linear transformations of the coordinates x^r which leave the components of the metric tensor g_{rs} invariant and which do not interchange past and future. We shall henceforth consider only transformations which leave the origin $x^r = 0$ fixed, i.e. homogeneous linear transformations. Then

(2.08) $$x'^r = L_s{}^r x^s$$

is a Lorentz transformation if

(2.09) $$g_{mn} L_r{}^m L_s{}^n = g_{rs}, \quad L_0{}^0 > 0.$$

It immediately follows that the determinant of a Lorentz transformation is $+1$ or -1. Lorentz transformations with determinant $+1$ are called ***proper***.

We associate a real 4-vector A^r with a Hermitian spintensor $a^{\dot\alpha\beta}$ by the relations:

(2.10) $$\begin{aligned} a^{\dot 1 1} &= A^0 + A^3, & A^0 &= \tfrac{1}{2}(a^{\dot 1 1} + a^{\dot 2 2}),\\ a^{\dot 1 2} &= A^1 - iA^2, & A^1 &= \tfrac{1}{2}(a^{\dot 1 2} + a^{\dot 2 1}),\\ a^{\dot 2 1} &= A^1 + iA^2, & A^2 &= \tfrac{1}{2}i(a^{\dot 1 2} - a^{\dot 2 1}),\\ a^{\dot 2 2} &= A^0 - A^3, & A^3 &= \tfrac{1}{2}(a^{\dot 1 1} - a^{\dot 2 2}). \end{aligned}$$

We then have

(2.11) $$g_{mn} A^m A^n = a^{\dot 1 1} a^{\dot 2 2} - a^{\dot 1 2} a^{\dot 2 1} = \det(a^{\dot\alpha\beta}).$$

From this identity it follows that spin-transformations $\lambda_\beta{}^\alpha$, which leave the determinant of an arbitrary Hermitian spintensor $a^{\dot\alpha\beta}$ invariant, induce transformations $L_s{}^r$ of Minkowski space-time which leave $g_{mn} A^m A^n$ invariant, i.e. Lorentz transformations. Now

$$\det(a'^{\dot\alpha\beta}) = \det(a^{\dot\sigma\rho}\, \lambda_{\dot\sigma}{}^{\dot\alpha}\, \lambda_\rho{}^\beta) = \det(a^{\dot\alpha\beta})\, |\det(\lambda_\beta{}^\alpha)|^2.$$

Thus we obtain the result: A spin-transformation $\lambda_\beta{}^\alpha$ induces a Lorentz transformation if and only if the absolute value of its determinant is unity, i.e.

(2.12) $$|\det(\lambda_\beta{}^\alpha)| = 1.$$

It is easily seen that the two spin-transformations $\lambda_\beta{}^\alpha$ and $\lambda_\beta{}^\alpha e^{i\theta}$ (θ any real number) induce the same Lorentz transformation. It follows that we may limit the spin-transformations to those with determinant $+1$, without reducing the set of Lorentz transformations which are induced by them. It can also be shown[9] that every proper Lorentz transformation can be obtained from a spin-transformation. We summarize our conclusions as follows:

A proper Lorentz transformation determines a spin-transformation, which satisfies (2.12), uniquely to within an arbitrary *phase factor* $e^{i\theta}$.

Every spin-transformation $\lambda_\beta{}^\alpha$, satisfying

(2.13) $$\det(\lambda_\beta{}^\alpha) = 1,$$

induces a proper Lorentz transformation. To every proper Lorentz

---

[9] O. Veblen and J. von Neumann, "Geometry of Complex Domains," Institute for Advanced Study mimeographed notes (Princeton, 1936).

transformation there correspond exactly two spin-transformations which satisfy (2.13) and which differ in sign only.

Spin-transformations which satisfy (2.13) leave invariant the components of the real skew-symmetric spintensor $\epsilon^{\alpha\beta}$, defined by:

$$\epsilon^{11} = \epsilon^{22} = 0, \quad \epsilon^{12} = -\epsilon^{21} = 1, \quad \epsilon^{\dot\alpha\dot\beta} = \epsilon^{\alpha\beta}. \tag{2.14}$$

This spintensor may be used to lower the suffixes of other spinors, and thus to introduce *covariant spinors* c_α, $a_{\dot\alpha\beta}$, etc. as follows:

$$\begin{aligned} c^\alpha &= \epsilon^{\alpha\beta} c_\beta, \\ c^1 &= c_2, \quad c^2 = -c_1; \end{aligned} \tag{2.15}$$

$$a^{\dot\alpha\beta} = \epsilon^{\dot\alpha\dot\gamma}\epsilon^{\beta\delta} a_{\dot\gamma\delta}, \tag{2.16}$$

$$a^{\dot11} = a_{\dot22}, \quad a^{\dot12} = -a_{\dot21}, \quad a^{\dot21} = -a_{\dot12}, \quad a^{\dot22} = a_{\dot11}.$$

In particular, we find that

$$\epsilon_{11} = \epsilon_{22} = 0, \quad \epsilon_{12} = -\epsilon_{21} = 1. \tag{2.17}$$

If the vector A^r is associated with the Hermitian spintensor $a^{\dot\alpha\beta}$ by the relations (2.10), and the vector B^r associated similarly with the spintensor $b^{\dot\alpha\beta}$, we deduce easily that

$$A^m B_m = g_{mn} A^m B^n = \tfrac{1}{2} a^{\dot\alpha\beta} b_{\dot\alpha\beta}. \tag{2.18}$$

3. The Cubic Lattice, Integral Null Vectors, Integral Spinvectors. Consider the points in Minkowski space whose coordinates t, x, y, z are all real integers. The set of these points will be called *the cubic lattice*.

The coordinates of a point P of the lattice may be regarded as the components of the vector OP which joins the origin O to P. Such a vector will be called an *integral vector* since its components are integers. For most purposes it suffices to restrict ourselves to *primitive integral vectors*, whose components have no common factor, as all other integral vectors are multiples of these.

As in the previous section, we can associate with an integral vector t, x, y, z a Hermitian spintensor $a^{\dot\alpha\beta}$ by the relations

$$\begin{aligned} a^{\dot11} &= t + z, \quad & a^{\dot12} &= x - iy, \\ a^{\dot21} &= x + iy, \quad & a^{\dot22} &= t - z, \end{aligned} \tag{3.01}$$

and

$$\begin{aligned} t &= \tfrac{1}{2}(a^{\dot11} + a^{\dot22}), \\ x &= \tfrac{1}{2}(a^{\dot12} + a^{\dot21}), \\ y &= \tfrac{1}{2}i(a^{\dot12} - a^{\dot21}), \\ z &= \tfrac{1}{2}(a^{\dot11} - a^{\dot22}). \end{aligned} \tag{3.02}$$

It immediately follows from (3.01) that the components of $a^{\dot\alpha\beta}$ are Gaussian integers. We also see from (3.01) that any common factor of t, x, y, z must be a factor of all $a^{\dot\alpha\beta}$. Equation (3.02) shows that any factor common to the $a^{\dot\alpha\beta}$, other than a factor of 2, must be a factor of t, x, y, z. In particular, if the vector t, x, y, z is to be primitive, any common factor of $a^{\dot\alpha\beta}$ must be a factor of 2.

We shall now study ***integral null vectors***, whose components satisfy the equation

$$t^2 - x^2 - y^2 - z^2 = 0. \tag{3.03}$$

By (2.11), this implies

$$a^{\dot{1}1}\, a^{\dot{2}2} = a^{\dot{1}2}\, a^{\dot{2}1}\,. \tag{3.04}$$

Making use of the unique factorization theorem for Gaussian integers, $a^{\dot{1}1}$ must split into two factors, of which one is a factor of $a^{\dot{1}2}$, the other being a factor of $a^{\dot{2}1}$. Since $a^{\dot{1}1}$ is real, those factors can be written in the form $mc^{\dot{1}}$ and nc^1, where m and n are real and relatively prime, and where $c^{\dot{1}} = \overline{c^1}$. Similarly, $a^{\dot{2}2}$ splits into factors $rc^{\dot{2}}$ and sc^2, $rc^{\dot{2}}$ being a factor of $a^{\dot{2}1}$ and sc^2 a factor of $a^{\dot{1}2}$, where r and s are real and relatively prime, and where $c^{\dot{2}} = \overline{c^2}$. Thus we have

$$a^{\dot{1}1} = mn\, c^{\dot{1}}c^1\,, \quad a^{\dot{1}2} = ms\, c^{\dot{1}}c^2\,,$$
$$a^{\dot{2}1} = rn\, c^{\dot{2}}c^1\,, \quad a^{\dot{2}2} = rs\, c^{\dot{2}}c^2\,.$$

Since $a^{\dot{1}2} = \overline{a^{\dot{2}1}}$, we have $ms = rn$. It follows that $m = r$, $n = s$, or $m = -r$, $n = -s$. Then

$$a^{\dot{1}1} = \quad mn\, c^{\dot{1}}c^1\,, \quad a^{\dot{1}2} = \pm\, mn\, c^{\dot{1}}c^2\,,$$
$$a^{\dot{2}1} = \pm\, mn\, c^{\dot{2}}c^1\,, \quad a^{\dot{2}2} = \quad mn\, c^{\dot{2}}c^2\,.$$

The factor ± 1 in the second and third of these expressions can be removed by absorbing it in c^1 or in c^2. Doing this and writing p for mn, we have

$$a^{\dot{\alpha}\beta} = pc^{\dot{\alpha}}c^{\beta}, \tag{3.05}$$

where p is a real integer. Decomposing p in the form (1.04), i.e. $p = qa\bar{a}$, we can absorb the complex integer a in both c^1 and c^2, thus reducing (3.05) to the form

$$a^{\dot{\alpha}\beta} = qc^{\dot{\alpha}}c^{\beta}, \tag{3.06}$$

where, as is easily seen, q is the product of those real primes, each taken once, which are contained an odd number of times in the greatest common factor of t, x, y, z.

Let us now consider ***primitive integral null vectors***. Since the square of a real integer leaves a remainder of 1 or 0 on division by 4, according as the integer is odd or even, it is easily seen from (3.03) that, of the components of a primitive integral null vector, t and one of x, y, z must be odd, while the two remaining components (two of x, y, z) must be even.

For a primitive integral null vector, q in (3.06) must be ± 1, and we arrive at the following result:

Each primitive integral null vector determines a spinvector c^{α} with integral components c^1, c^2, such that

$$a^{\dot{\alpha}\beta} = \pm\, c^{\dot{\alpha}}c^{\beta}, \tag{3.07}$$

where the upper or lower sign must be taken throughout. Since $c^{\dot{1}}c^1$ and $c^{\dot{2}}c^2$ are both positive, we see from (3.02) that t is positive or negative according as the plus or minus sign is chosen in (3.07). For primitive integral null vectors pointing into the future we have $t > 0$, and thus

$$a^{\dot{\alpha}\beta} = c^{\dot{\alpha}}c^{\beta}. \tag{3.08}$$

Some non-primitive null vectors pointing into the future can also be represented in the form (3.08). Whether this is possible or whether the representation takes the more complicated form (3.06), with $q > 1$, depends only on the properties of the greatest common factor of the components of the null vector.

From (3.08) and (3.02) it is seen that a spinvector c^α with integral components determines an integral null vector (t, x, y, z) if and only if *c^1, c^2 are both odd or both even.* Such a spinvector will be called an ***integral spinvector***. Note that, even if c^1, c^2 are integers, c^α is not an integral spinvector if $c^1 + c^2$ is odd.

The sum and difference of integral spinvectors are again integral spinvectors; the product of an integer and an integral spinvector is an integral spinvector. Thus the integral spinvectors form a two-dimensional complex vector space with coefficients in the ring of complex integers. It is easy to see that the independent integral spinvectors

$$\mathbf{e}_{(1)} \equiv (1 + i\,, 0), \quad \mathbf{e}_{(2)} \equiv (1\,, 1) \tag{3.09}$$

form a basis; this means that any integral spinvector $\mathbf{c}$ can be written in the form

$$\mathbf{c} = a\mathbf{e}_{(1)} + b\mathbf{e}_{(2)}\,, \tag{3.10}$$

where a and b are integers, and that conversely any spinvector of this form is integral.

The following theorem can be derived:

The null vector associated, by (3.08), (3.02), *with an integral spinvector c^α is integral and primitive if and only if one or other of the following two conditions is satisfied:*

(3.11)

I c^1, c^2 *are both odd and relatively prime.*

II $c^\alpha = (1 + i)d^\alpha$,
where d^1, d^2 are relatively prime and one of them is even, the other odd.

In the first case t is odd, z is even, and one of x, y is odd, the other even; in the second case t, z are odd and x, y are even. By (3.08), $\pm c^\alpha$ and $\pm\, ic^\alpha$ determine the same null vector.

The criterion which we have just stated solves our basic problem of determining all primitive integral null vectors.

If we drop the requirement that c^1, c^2 be integers, we may enquire to what extent the spinvector c^α is determined by a null vector in space-time. By (3.01), a null vector determines a unique Hermitian spintensor $a^{\dot\alpha\beta}$. Let

$$a^{\dot\alpha\beta} = c^{\dot\alpha}\, c^{\beta} = c'^{\dot\alpha}\, c'^{\beta}\,, \tag{3.12}$$

or, equivalently,

$$c^{\dot 1}c^1 = c'^{\dot 1}c'^1 \quad , \quad c^{\dot 2}c^2 = c'^{\dot 2}c'^2\,,$$
$$c^{\dot 1}c^2 = c'^{\dot 1}c'^2 \quad , \quad c^{\dot 2}c^1 = c'^{\dot 2}c'^1\,.$$

The first two of these equations show that $c'^1 = c^1 e^{i\theta}$, $c'^2 = c^2 e^{i\phi}$ (θ, ϕ real), and the last two equations imply $\theta = \phi$. Hence

$$c'^\alpha = c^\alpha e^{i\theta}. \tag{3.13}$$

Thus a null vector t, x, y, z, determines a spinvector c^α uniquely to within an arbitrary phase factor $e^{i\theta}$.

4. Integral Null Vectors are Spatially Dense. In (3.08) let us write

$$c^1 = (1 + i)(p + iq) \quad , \quad c^2 = (1 + i)r, \tag{4.01}$$

where p, q, r are real integers. By (3.02) the spinvector c^α determines a null vector with components

$$t = p^2 + q^2 + r^2, \tag{4.02}$$

$$x = 2pr, \quad y = 2qr, \quad z = p^2 + q^2 - r^2. \tag{4.03}$$

Equations (4.02), (4.03) determine a discrete three parameter set of integral null vectors which are not necessarily primitive. We shall now show that the spatial projections of these null vectors, i.e. the directions defined by (4.03), are everywhere dense.

Consider an arbitrary direction D_0 in space, and let l_0, m_0, n_0 be its direction cosines. We can then define real numbers a_0, b_0, c_0 by the equations

$$l_0 = 2a_0c_0, \quad m_0 = 2b_0c_0, \quad n_0 = a_0^2 + b_0^2 - c_0^2. \tag{4.04}$$

We obtain, by virtue of $l_0^2 + m_0^2 + n_0^2 = 1$,

$$a_0 = l_0[2(1 - n_0)]^{-\frac{1}{2}}, \quad b_0 = m_0[2(1 - n_0)]^{-\frac{1}{2}}, \quad c_0 = [\tfrac{1}{2}(1 - n_0)]^{\frac{1}{2}}. \tag{4.05}$$

It is obvious that a_0, b_0, c_0 can be approximated by *rational* fractions a, b, c such that l, m, n, defined by[10]

$$l = 2ac, \quad m = 2bc, \quad n = a^2 + b^2 - c^2, \tag{4.06}$$

are arbitrarily close to l_0, m_0, n_0, respectively. Thus the direction D, with direction ratios l, m, n, makes an arbitrarily small angle with D_0. Let the integer d be the least common denominator of the rational fractions a, b, c. Then p, q, r, defined by

$$p = ad, \quad q = bd, \quad r = cd, \tag{4.07}$$

are real integers. If we substitute these integers into (4.02) and (4.03) we obtain an integral null line whose spatial component (x, y, z) is immediately seen to have the direction D. Since D approximates D_0, our assertion is proved.

Having shown that a subset of all integral null vectors is spatially dense, it follows, *a fortiori*, that the same is true for the set of all integral null vectors. Since every integral null vector is codirectional with a primitive integral null vector, the *set of all primitive integral null vectors is spatially dense.*[11]

---

[10] $l^2 + m^2 + n^2$ is not necessarily 1.

[11] By saying that a set of vectors is spatially dense, we mean, more precisely, that the *directions* of the spatial projections of the vectors in the set are dense. This remark applies also to Sec. 8.

5. Integral Lorentz Transformations. A Lorentz transformation

$$x'^r = L_s{}^r x^s \tag{5.01}$$

is ***integral*** if it maps into itself, i.e. leaves invariant as a whole, the cubic lattice which consists of all points with integral coordinates x^r.

Consider the integral vector $x^s \equiv \delta_m{}^s$, where m is 0, 1, 2, or 3, and where $\delta_m{}^s$ is 1 if $s = m$ and 0 if $s \neq m$. The transformation (5.01) maps this vector into $x'^r \equiv L_m{}^r$. If (5.01) is an integral Lorentz transformation then x'^r must be an integral vector; thus the components $L_m{}^r$ must be real integers. It is obvious that then x'^r is always integral whenever x^r is.

Since the determinant of a Lorentz transformation is ± 1, the components $(L^{-1})_s{}^r$ of the inverse Lorentz transformation will be real integers if $L_s{}^r$ are real integers. Thus a Lorentz transformation with integral components $L_s{}^r$ maps the set of all integral vectors into the set of all integral vectors, and not into a proper subset of the latter. The following conclusion is immediate:

A Lorentz transformation $L_s{}^r$ is integral if and only if all its components $L_s{}^r$ are real integers.

Thus, by (2.09), our problem of determining all integral Lorentz transformations reduces to the solution of 10 quadratic diophantine equations in 16 unknown integers. This rather formidable mathematical problem can be approached indirectly by considering integral null vectors, spinvectors and spin-transformations, as will be shown in this section and the next.

We shall now prove the following theorem: A necessary and sufficient condition for a Lorentz transformation to be integral is that the Lorentz transformation, as well as its inverse, map primitive integral null vectors into integral null vectors. The necessity of the condition is trivial; we shall therefore consider only its sufficiency.

Consider the four independent primitive integral null vectors

$$\begin{aligned} \mathbf{N}_{(0)} &\equiv (1, -1, 0, 0), \\ \mathbf{N}_{(1)} &\equiv (1, 1, 0, 0), \\ \mathbf{N}_{(2)} &\equiv (1, 0, 1, 0), \\ \mathbf{N}_{(3)} &\equiv (1, 0, 0, 1). \end{aligned} \tag{5.02}$$

We have

$$\begin{aligned} \mathbf{E}_{(0)} &\equiv (1, 0, 0, 0) = \tfrac{1}{2}(\mathbf{N}_{(0)} + \mathbf{N}_{(1)}), \\ \mathbf{E}_{(1)} &\equiv (0, 1, 0, 0) = \tfrac{1}{2}(-\mathbf{N}_{(0)} + \mathbf{N}_{(1)}), \\ \mathbf{E}_{(2)} &\equiv (0, 0, 1, 0) = \mathbf{N}_{(2)} - \tfrac{1}{2}(\mathbf{N}_{(0)} + \mathbf{N}_{(1)}), \\ \mathbf{E}_{(3)} &\equiv (0, 0, 0, 1) = \mathbf{N}_{(3)} - \tfrac{1}{2}(\mathbf{N}_{(0)} + \mathbf{N}_{(1)}). \end{aligned} \tag{5.03}$$

A Lorentz transformation, satisfying the hypothesis of our theorem, maps the vectors $\mathbf{N}_{(r)}$ into integral null vectors

$$\mathbf{N}'_{(r)} \equiv (t_{(r)}, x_{(r)}, y_{(r)}, z_{(r)}), \quad r = 0, 1, 2, 3, \tag{5.04}$$

and it maps the vectors $\mathbf{E}_{(r)}$ into

$$\begin{aligned} \mathbf{E}'_{(0)} &= \tfrac{1}{2}(\mathbf{N}'_{(0)} + \mathbf{N}'_{(1)}), \\ \mathbf{E}'_{(1)} &= \tfrac{1}{2}(-\mathbf{N}'_{(0)} + \mathbf{N}'_{(1)}), \\ \mathbf{E}'_{(2)} &= \mathbf{N}'_{(2)} - \tfrac{1}{2}(\mathbf{N}'_{(0)} + \mathbf{N}'_{(1)}), \\ \mathbf{E}'_{(3)} &= \mathbf{N}'_{(3)} - \tfrac{1}{2}(\mathbf{N}'_{(0)} + \mathbf{N}'_{(1)}). \end{aligned} \tag{5.05}$$

If $\mathbf{N'}_{(r)} = \mathbf{M'}_{(r)}d$, where d is an integer and $\mathbf{M'}_{(r)}$ an integral null vector, then, by hypothesis, the inverse of the Lorentz transformation considered here maps $\mathbf{M'}_{(r)}$ into an integral null vector $\mathbf{M}_{(r)}$, and therefore maps $\mathbf{N'}_{(r)}$ into $\mathbf{N}_{(r)} = \mathbf{M}_{(r)}d$. But, by (5.02), the vectors $\mathbf{N}_{(r)}$ are all primitive. It follows that d must be a unit. Hence the integral null vectors $\mathbf{N'}_{(r)}$ must be primitive. Then by Sec. **3**, $t_{(r)}$ and one of $x_{(r)}$, $y_{(r)}$, $z_{(r)}$ must be odd, the other two being even.

Since the scalar product of two vectors is an invariant, we have

$$g_{mn}\, \mathbf{N'}_{(0)}{}^m\, \mathbf{N'}_{(1)}{}^n \;=\; g_{mn}\, \mathbf{N}_{(0)}{}^m\, \mathbf{N}_{(1)}{}^n,$$

or

(5.06) $$t_{(0)}\, t_{(1)} - x_{(0)}\, x_{(1)} - y_{(0)}\, y_{(1)} - z_{(0)}\, z_{(1)} = 2.$$

Thus the left-hand side of this equation is even. Combining this fact with the last statement of the preceding paragraph, we see that $t_{(0)}\, t_{(1)}$ and one of the three products $x_{(0)}\, x_{(1)}$, $y_{(0)}\, y_{(1)}$, $z_{(0)}\, z_{(1)}$ must be odd. In order to be definite, let us take $x_{(0)}\, x_{(1)}$ odd. Then $t_{(0)}$, $t_{(1)}$, $x_{(0)}$, $x_{(1)}$ are odd, and $y_{(0)}$, $y_{(1)}$, $z_{(0)}$, $z_{(1)}$ are even. It follows that $t_{(1)} \pm t_{(0)}$, $x_{(1)} \pm x_{(0)}$, $y_{(1)} \pm y_{(0)}$, $z_{(1)} \pm z_{(0)}$ are all even integers. Hence $\mathbf{E'}_{(r)}$, defined in (5.05), are integral vectors.

Applying the Lorentz transformation (5.01) to the vectors $\mathbf{E}_{(r)}$, given by (5.03), we obtain

(5.07) $$\mathbf{E'}_{(r)}{}^s = L_r{}^s\,.$$

Thus $L_r{}^s$ are integers and the sufficiency of our condition is demonstrated.

6. Integral Spin-Transformations. It seems legitimate to deduce from the preceding theorem that a spin-transformation is associated with an integral Lorentz transformation if and only if both the spin-transformation and its inverse map integral spinvectors into integral spinvectors. It must be pointed out, though, that this statement is not *a priori* obvious and that we must proceed with caution. The reason is that a spinvector is not uniquely determined by a primitive null vector in space-time, but is determined only to within an arbitrary phase factor. However, we shall show now that the statement made above is true if the spin-transformation is taken with a suitable phase factor.

If a spin-transformation and its inverse map integral spinvectors into integral spinvectors, then the corresponding Lorentz transformation is integral since it and its inverse map primitive integral null vectors into integral null vectors. It is therefore sufficient to show that, given an integral Lorentz transformation, a spin-transformation can be found which represents it and which, as well as its inverse, maps integral spinvectors into integral spinvectors.

Let $L_s{}^r$ be an arbitrary integral Lorentz transformation. Since $L_s{}^r$ are integers, it is clear, by (5.01), that the greatest common factor of the components of an integral vector x^r is a common factor of the components of the transform x'^r of x^r under the integral Lorentz transformation. Since x^r can also be obtained from x'^r by the integral Lorentz transformation

$(L^{-1})_s{}^r$, it follows that the x^r and the x'^r have the same greatest common factor. In particular, we have that, if an integral null vector can be represented in the form (3.08), the transform of this null vector under any integral Lorentz transformation can again be so represented. The following is easily deduced:

If $\lambda_\beta{}^\alpha$ is a spin-transformation which represents an integral Lorentz transformation $L_s{}^r$ then $\lambda_\beta{}^\alpha$ maps any integral spinvector into a spinvector which differs from an integral spinvector by at most a phase factor.

Therefore, if we introduce the two spinvectors

(6.01) $$\mathbf{e}_{(1)} = (1 + i, 0), \quad \mathbf{e}_{(2)} = (1, 1),$$

we must have

(6.02) $$\lambda_\beta{}^\alpha e_{(1)}{}^\beta = e^{-i\phi} e'_{(1)}{}^\alpha, \qquad \lambda_\beta{}^\alpha e_{(2)}{}^\beta = e^{-i\theta} e'_{(2)}{}^\alpha,$$

where $\mathbf{e}'_{(1)}$, $\mathbf{e}'_{(2)}$, are integral spinvectors. Since $\lambda_\beta{}^\alpha$ is determined by $L_s{}^r$ only to within an arbitrary phase factor, we can choose this phase factor so that, in (6.02), $\phi = 0$. We can then write:

(6.03) $$(\lambda^{-1})_\beta{}^\alpha e'_{(1)}{}^\beta = e_{(1)}{}^\alpha, \qquad (\lambda^{-1})_\beta{}^\alpha e'_{(2)}{}^\beta = e^{i\theta} e_{(2)}{}^\alpha,$$

where $(\lambda^{-1})_\beta{}^\alpha$ is the inverse spin-transformation which exists, by (2.12), and which represents the integral Lorentz transformation $(L^{-1})_s{}^r$.

From (6.03) we obtain, on addition,

(6.04) $$(\lambda^{-1})_\beta{}^\alpha (e'_{(1)}{}^\beta + e'_{(2)}{}^\beta) = e_{(1)}{}^\alpha + e^{i\theta} e_{(2)}{}^\alpha.$$

Since $e'_{(1)}{}^\beta + e'_{(2)}{}^\beta$ is integral, $e_{(1)}{}^\alpha + e^{i\theta} e_{(2)}{}^\alpha$ must be of the form $e^{i\psi}(p, q)$, p and q being integers and ψ real. Thus, by (6.01), we have

(6.05) $$1 + i + e^{i\theta} = e^{i\psi} p,$$

(6.06) $$e^{i\theta} = e^{i\psi} q.$$

From (6.06) it follows that $|q| = 1$ and hence that q and $u = 1/q$ are units. Then (6.05) can be written

(6.07) $$1 + i = e^{i\theta}(up - 1).$$

Taking absolute values, we find that $|up - 1| = 2^{\frac{1}{2}}$ and thus $up - 1$ must be one of $1 + i$, $1 - i$, $-1 + i$, or $-1 - i$, since these are the only integers of absolute value $2^{\frac{1}{2}}$. In either of these cases $e^{-i\theta}$ is a unit, by (6.07). Then $e''_{(2)}{}^\beta = e^{-i\theta} e'_{(2)}{}^\beta$ is an integer and we can rewrite (6.03) as follows:

(6.08) $$\lambda_\beta{}^\alpha e_{(1)}{}^\beta = e''_{(1)}{}^\alpha, \qquad \lambda_\beta{}^\alpha e_{(2)}{}^\beta = e''_{(2)}{}^\alpha,$$

where $e''_{(1)}{}^\alpha = e'_{(1)}{}^\alpha$. Since an arbitrary integral spinvector can be written in the form (3.10), we immediately see that $\lambda_\beta{}^\alpha$ maps integral spinvectors into integral spinvectors. Similarly, corresponding to $(L^{-1})_s{}^r$, there must exist a spin-transformation $\mu_\beta{}^\alpha$ which maps integral spinvectors into integral spinvectors. But $\mu_\beta{}^\alpha$ can differ from $(\lambda^{-1})_\beta{}^\alpha$ by at most a phase factor $e^{i\chi}$ and, since $(\lambda^{-1})_\beta{}^\alpha$ maps $\mathbf{e}''_{(2)}$ into $\mathbf{e}_{(2)}$, $\mu_\beta{}^\alpha$ maps $\mathbf{e}''_{(2)}$ into an integral spinvector $e^{i\chi}\,\mathbf{e}_{(2)}$. It follows that $e^{i\chi}$ is a unit and that therefore $(\lambda^{-1})_\beta{}^\alpha$ maps integral spinvectors into integral spinvectors. This establishes our assertion.

Let us denote by v the determinant of $\lambda_\beta{}^\alpha$; we know, by (2.12), that $|v| = 1$. Writing (6.08) in the form

(6.09) $$\lambda_\beta{}^\alpha e_{(\gamma)}{}^\beta = e''_{(\gamma)}{}^\alpha,$$

and taking determinants on both sides, we have

$$v(1 + i) = \det(e''_{(\gamma)}{}^{\alpha}), \tag{6.10}$$

by (6.01). The right-hand side of (6.10) is obviously an integer. By the same argument as that applied to (6.07), it follows that v is a unit, i.e. $v = 1$, $v = i$, or $v = -1$, $v = -i$. In the latter two cases we can absorb the phase factor i in $\lambda_\beta{}^\alpha$, thus reducing these cases to the first two.

It is now clear that every proper integral Lorentz transformation is represented by two spin-transformations, differing in sign only, which satisfy the condition

$$\det(\lambda_\beta{}^\alpha) = 1 \text{ or } i, \tag{6.11}$$

and which are such that both the spin-transformation and its inverse map the two integral spinvectors $\mathbf{e}_{(1)}$ and $\mathbf{e}_{(2)}$, given by (6.01), into integral spinvectors. Conversely, the conditions just imposed on a spin-transformation are sufficient to insure that the spin-transformation corresponds to an integral Lorentz transformation.

Spin-transformations which satisfy the above conditions will be called *integral spin-transformations*. We shall now obtain the conditions on integral spin-transformations in a more explicit form.

If $\det(\lambda_\beta{}^\alpha) = 1$, (6.11), we have

$$(\lambda^{-1})_1{}^1 = \lambda_2{}^2, \; (\lambda^{-1})_2{}^1 = -\lambda_2{}^1, \; (\lambda^{-1})_1{}^2 = -\lambda_1{}^2, \; (\lambda^{-1})_2{}^2 = \lambda_1{}^1. \tag{6.12}$$

On transforming $\mathbf{e}_{(1)}$ and $\mathbf{e}_{(2)}$ by $\lambda_\beta{}^\alpha$ and by $(\lambda^{-1})_\beta{}^\alpha$ we obtain the following four spinvectors:

$$\begin{aligned} &((1+i)\lambda_1{}^1, (1+i)\lambda_1{}^2), \; (\lambda_1{}^1 + \lambda_2{}^1, \lambda_1{}^2 + \lambda_2{}^2), \\ &((1+i)\lambda_2{}^2, -(1+i)\lambda_1{}^2), \; (\lambda_2{}^2 - \lambda_2{}^1, -\lambda_1{}^2 + \lambda_1{}^1). \end{aligned} \tag{6.13}$$

If $\det(\lambda_\beta{}^\alpha) = i$, (6.11), we obtain by the same procedure four spinvectors which differ from those in (6.13) by unit factors only. Thus, in either case, each of the spinvectors (6.13) must be integral, i.e. the two components must be integers, both odd or both even. We easily deduce the following result:

A spin-transformation $\lambda_\beta{}^\alpha$ is integral if and only if one of the following four conditions is satisfied:

I *$\lambda_\beta{}^\alpha$ are integers such that*

$$\lambda_1{}^1\lambda_2{}^2 - \lambda_2{}^1\lambda_1{}^2 = 1, \tag{6.14}$$

and such that $\lambda_1{}^1 + \lambda_2{}^1 + \lambda_1{}^2 + \lambda_2{}^2$ is even.

II

$$\lambda_\beta{}^\alpha = \mu_\beta{}^\alpha/(1+i), \tag{6.15}$$

where $\mu_\beta{}^\alpha$ are odd integers such that

$$\mu_1{}^1\mu_2{}^2 - \mu_2{}^1\mu_1{}^2 = 2i. \tag{6.16}$$

III *$\lambda_\beta{}^\alpha$ are integers such that*

$$\lambda_1{}^1\lambda_2{}^2 - \lambda_2{}^1\lambda_1{}^2 = i, \tag{6.17}$$

and such that $\lambda_1{}^1 + \lambda_2{}^1 + \lambda_1{}^2 + \lambda_2{}^2$ is even.

IV $$\lambda_\beta{}^\alpha = \mu_\beta{}^\alpha/(1+i),$$

where $\mu_\beta{}^\alpha$ are odd integers such that

(6.18) $$\mu_1{}^1\mu_2{}^2 - \mu_2{}^1\mu_1{}^2 = -2.$$

In cases II and IV, (6.16) and (6.18) are, by (6.15), equivalent respectively to (6.14) and (6.17). In these cases the condition that the sum of the $\lambda_\beta{}^\alpha$ be an even integer need not be stated separately since it follows from the other requirements, as can be seen by examining the possible remainders of $\mu_\beta{}^\alpha$ on division by 2.

The integral spin-transformations of types III and IV can be replaced by spin-transformations of determinant $+1$ if the phase factor $e^{-i\pi/4} = 2^{\frac{1}{2}}/(1+i)$ is introduced. This procedure has the disadvantage of introducing the irrationality $2^{\frac{1}{2}}$, but it has the advantage that the resulting spin-transformations together with those of types I and II form a group.

From the discussion of the diophantine equation (6.14) in Sec. 1 we see that there is a discrete sixfold infinity of integral spin-transformations of the type I. Similarly, it can be shown for each of the types II, III, and IV, that there is a discrete sixfold infinity of integral spin-transformations. Since there is a 2 - 1 correspondence between integral spin-transformations and proper integral Lorentz transformations, we have:

The group of proper integral Lorentz transformations is a discrete, sixfold infinite set.

We have not hesitated to "count" the order of infinity of the integral Lorentz group because this emphasizes the large number of integral Lorentz transformations. However, since we are dealing with an enumerably infinite discrete group of transformations without infinitesimal elements, the statement that the group is sixfold infinite has no invariant significance and must not be taken too literally. A different parametrization of the elements of the group may easily result in an order of infinity other than six.

7. Equivalence of Primitive Integral Null Vectors. We shall now prove that, *given two primitive integral null vectors, an integral Lorentz transformation can be found which maps the one into the other.* Thus all primitive integral null vectors are equivalent in the sense that no single such vector possesses an invariant property which is not shared by all others.

In Sec. 3 we saw that if the vector (t, x, y, z) is a primitive integral null vector, then t is odd and one of x, y, z is odd, the remaining two components being even. A primitive integral null vector with y or x odd is mapped into a vector with z odd by the proper integral Lorentz transformation which cyclicly permutes the $x-$, $y-$, and $z-$ axes once or twice. It follows that it is sufficient to prove the italicized statement for the case where the two assigned primitive integral null vectors have odd $z-$ components. Then, by Sec. 3, the two null vectors are represented by spinvectors of the form (3.11):

(7.01) $$c^1 = (1+i)d^1, \quad c^2 = (1+i)d^2,$$

where d^1 and d^2 are relatively prime integers of which one is even and one

odd. It is obviously sufficient to show that there always exists an integral spin-transformation mapping an integral spinvector of the type considered into the spinvector $\mathbf{e}_{(1)} = (1 + i, 0)$.

Consider the spin-transformation

$$\lambda_\beta{}^\alpha\, c^\beta = e_{(1)}{}^\alpha. \tag{7.02}$$

By (7.01), we can write this

$$\lambda_1{}^1\, d^1 + \lambda_2{}^1\, d^2 = 1, \tag{7.03}$$

$$\lambda_1{}^2\, d^1 + \lambda_2{}^2\, d^2 = 0. \tag{7.04}$$

The last equation is satisfied if we put

$$\lambda_1{}^2 = -\,d^2, \quad \lambda_2{}^2 = d^1. \tag{7.05}$$

Then equation (7.03) is identical with the condition (6.14). Since, by (7.05), $\lambda_1{}^2$ and $\lambda_2{}^2$ are relatively prime integers and $\lambda_1{}^2 + \lambda_2{}^2$ is odd, we can, by Sec. 1, find integers $\lambda_1{}^1$ and $\lambda_2{}^1$ satisfying (7.03), or equivalently (6.14), and such that $\lambda_1{}^1 + \lambda_2{}^1$ is odd. It follows that $\lambda_1{}^1 + \lambda_2{}^1 + \lambda_1{}^2 + \lambda_2{}^2$ is even. Thus conditions I (Sec. 6) for integral spin-transformations are satisfied by $\lambda_\beta{}^\alpha$ and our proof is complete.

8. Integral Time Lines are Spatially Dense. *Integral time lines* are the transforms of the t-axis ($x = y = z = 0$) under integral Lorentz transformations. A primitive integral vector having the direction of an integral time line will be called a ***primitive integral time vector***; it is the transform under an integral Lorentz transformation of the vector

$$\mathbf{E}_{(0)} = (1, 0, 0, 0). \tag{8.01}$$

Thus far integral Lorentz transformations have been regarded as mappings of space-time into itself, which map the points of the cubic lattice into other lattice points. However, an integral Lorentz transformation can also be regarded as a change to a new coordinate system, such that the points of the cubic lattice have again integral coordinates with respect to the new coordinate axes. Such a coordinate system will be called an ***integral Lorentz frame***. Integral time lines are merely the t-axes of integral Lorentz frames.

Before we investigate the main theorem of this section we shall consider briefly the velocities associated with integral time lines. By "velocity" is meant the velocity of a particle whose world line coincides with the integral time line, or, equivalently, the velocity of a particle at rest in the corresponding integral Lorentz frame.

The components t, x, y, z of a primitive integral time vector satisfy the diophantine equation

$$t^2 - x^2 - y^2 - z^2 = 1. \tag{8.02}$$

The velocity v associated with this integral time vector is given by

$$v = \left(\frac{x^2}{t^2} + \frac{y^2}{t^2} + \frac{z^2}{t^2}\right)^{\frac{1}{2}}. \tag{8.03}$$

By (8.02), this reduces to

$$v = \left(1 - \frac{1}{t^2}\right)^{\frac{1}{2}} = \frac{1}{t}(t^2 - 1)^{\frac{1}{2}}. \tag{8.04}$$

Since t must be an integer we see that the only possible velocities are, for $t = 1, 2, 3, 4, \ldots,$

$$v = 0, \quad \tfrac{1}{2}\,3^{\frac{1}{2}}, \quad \tfrac{2}{3}\,2^{\frac{1}{2}}, \quad \tfrac{1}{4}\,(15)^{\frac{1}{2}}, \ldots. \tag{8.05}$$

Remembering that we have chosen the velocity of light c equal to unity, we see that *the velocities (other than zero) associated with integral time lines are very high, the smallest velocity being* $\tfrac{1}{2}3^{\frac{1}{2}} = 0.866$ *times the velocity of light.*

An example of an integral time line, associated with the minimum non-zero velocity $\tfrac{1}{2}3^{\frac{1}{2}}$, is given by the transform of the t-axis under the integral Lorentz transformation

$$L_s{}^r = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}. \tag{8.06}$$

We now proceed to show that integral time lines are spatially dense. The vector $\mathbf{E}_{(0)}$ (8.01) is associated with the Hermitian spintensor $a_{(0)}{}^{\dot\alpha\beta}$, given by

$$a_{(0)}{}^{\dot1 1} = a_{(0)}{}^{\dot2 2} = 1, \quad a_{(0)}{}^{\dot1 2} = a_{(0)}{}^{\dot2 1} = 0. \tag{8.07}$$

The integral spin-transformation $\lambda_\beta{}^\alpha$ of type I (Sec. 6) maps $a_{(0)}{}^{\dot\alpha\beta}$ into the spintensor $a^{\dot\alpha\beta}$, given by

$$\begin{aligned} a^{\dot1 1} &= \bar\lambda_1{}^1\,\lambda_1{}^1 + \bar\lambda_2{}^1\,\lambda_2{}^1, & a^{\dot1 2} &= \bar\lambda_1{}^1\,\lambda_1{}^2 + \bar\lambda_2{}^1\,\lambda_2{}^2, \\ a^{\dot2 1} &= \bar\lambda_1{}^2\,\lambda_1{}^1 + \bar\lambda_2{}^2\,\lambda_2{}^1, & a^{\dot2 2} &= \bar\lambda_1{}^2\,\lambda_1{}^2 + \bar\lambda_2{}^2\,\lambda_2{}^2, \end{aligned} \tag{8.08}$$

and $a^{\dot\alpha\beta}$ is in turn associated with the primitive integral time vector whose components t, x, y, z are given by

$$\begin{aligned} t &= \tfrac{1}{2}(\bar\lambda_1{}^1\,\lambda_1{}^1 + \bar\lambda_2{}^1\,\lambda_2{}^1 + \bar\lambda_1{}^2\,\lambda_1{}^2 + \bar\lambda_2{}^2\,\lambda_2{}^2), \\ x + iy &= \bar\lambda_1{}^2\,\lambda_1{}^1 + \bar\lambda_2{}^2\,\lambda_2{}^1, \\ z &= \tfrac{1}{2}(\bar\lambda_1{}^1\,\lambda_1{}^1 + \bar\lambda_2{}^1\,\lambda_2{}^1 - \bar\lambda_1{}^2\,\lambda_1{}^2 - \bar\lambda_2{}^2\,\lambda_2{}^2). \end{aligned} \tag{8.09}$$

We shall now show that the spatial projections of the integral time vectors (8.09) are dense, or, equivalently, that the expression

$$\frac{x}{z} + i\frac{y}{z} = \frac{2(\bar\lambda_1{}^2\,\lambda_1{}^1 + \bar\lambda_2{}^2\,\lambda_2{}^1)}{\bar\lambda_1{}^1\,\lambda_1{}^1 + \bar\lambda_2{}^1\,\lambda_2{}^1 - \bar\lambda_1{}^2\,\lambda_1{}^2 - \bar\lambda_2{}^2\,\lambda_2{}^2} \tag{8.10}$$

can be made to approximate to an arbitrary degree any preassigned complex number β, which we may assume to be non-zero.

Given an arbitrary non-zero complex number β, we define α by the equation

$$\beta = \frac{2\alpha}{\bar\alpha\alpha - 1}. \tag{8.11}$$

We may take α to be

$$\alpha = [1 + (1 + \bar\beta\beta)^{\frac{1}{2}}]/\bar\beta, \tag{8.12}$$

so that $|\alpha| > 1$, and therefore $\bar\alpha\alpha - 1 > 0$.

It is obvious that, given a small positive ϵ, we can find complex integers $\lambda_1{}^1$, and $\lambda_1{}^2$, which are relatively prime, such that

$$\left| \alpha - \frac{\lambda_1{}^1}{\lambda_1{}^2} \right| < \frac{\epsilon}{2}, \quad |\lambda_1{}^2| > \frac{2}{\epsilon}, \tag{8.13}$$

and such that $\lambda_1{}^1$ is even. Then, by Sec. 1, non-zero integers $\lambda_2{}^2$, $\lambda_2{}^1$, can be found, satisfying

$$\lambda_1{}^1 \lambda_2{}^2 - \lambda_2{}^1 \lambda_1{}^2 = 1, \tag{8.14}$$

and such that $\lambda_1{}^1 + \lambda_2{}^1 + \lambda_1{}^2 + \lambda_2{}^2$ is even. Then $\lambda_\beta{}^\alpha$ are the components of an integral spin-transformation of type I (Sec. 6). From (8.14) we obtain

$$\left| \frac{\lambda_1{}^1}{\lambda_1{}^2} - \frac{\lambda_2{}^1}{\lambda_2{}^2} \right| = \left| \frac{1}{\lambda_1{}^2 \lambda_2{}^2} \right| < \frac{\epsilon}{2}, \tag{8.15}$$

by the second inequality of (8.13). Combining (8.13) with (8.15), we have

$$\left| \alpha - \frac{\lambda_2{}^1}{\lambda_2{}^2} \right| < \epsilon. \tag{8.16}$$

Thus both $\lambda_1{}^1/\lambda_1{}^2$ and $\lambda_2{}^1/\lambda_2{}^2$ approximate α. Substituting in (8.11), we see that the number β is approximated by the two fractions

$$\frac{2\bar{\lambda}_1{}^2 \lambda_1{}^1}{\bar{\lambda}_1{}^1 \lambda_1{}^1 - \bar{\lambda}_1{}^2 \lambda_1{}^2} \quad \text{and} \quad \frac{2\bar{\lambda}_2{}^2 \lambda_2{}^1}{\bar{\lambda}_2{}^1 \lambda_2{}^1 - \bar{\lambda}_2{}^2 \lambda_2{}^2}, \tag{8.17}$$

the two denominators being positive, since $\bar{\alpha}\alpha - 1 > 0$. It follows that β is approximated by the fraction which is obtained by adding the numerators and denominators of the fractions (8.17), i.e. β is approximated by (8.10). This completes the proof that we can find integral time lines whose spatial projections approximate any preassigned direction in space.

Since every integral time line is codirectional with a primitive integral time vector, we deduce the following theorem:

The set of all primitive integral time vectors is spatially dense.

We add, without proof, the statement of a more general theorem which can be verified by arguments more complicated than, but quite similar to those just given above.

Consider an integral Lorentz frame and any integral vector. *The transforms of the integral vector under all integral Lorentz transformations form a set which is spatially dense.*

The preceding theorem is a special case of this. So also is the theorem of Sec. 4, once the equivalence of primitive null vectors (Sec. 7) is established.

The author wishes to thank Professors H. S. M. Coxeter, R. J. Duffin, R. P. Feynman, L. Infeld, and J. L. Synge for interesting discussions and helpful suggestions.

This work was completed last winter when the author was a Frank B. Jewett Fellow, resident at the Institute for Advanced Study, Princeton. It is a pleasure to thank the Frank B. Jewett Fellowship Committee for their award, and the Institute for its hospitality.

Appendix

Professor H. S. M. Coxeter was kind enough to show me some independent work of his which is essentially equivalent to our problem of finding all integral Lorentz transformations. He considers a lattice in *hyperbolic 3-space* consisting of the points of our cubic lattice which lie on the unit "sphere"

$$t^2 - x^2 - y^2 - z^2 = 1. \tag{A}$$

The congruent transformations of hyperbolic space which leave this lattice invariant as a whole are exactly our integral Lorentz transformations.

Coxeter chooses as his basic operation the reflection in 4-space which consists of adding the quantity $t - x - y - z$ to each of the four coordinates t, x, y, z of a point. In our notation this transformation is given by

$$L_s{}^r = \begin{pmatrix} 2 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & 0 \end{pmatrix}. \tag{B}$$

This is easily seen to be an integral Lorentz transformation. Combining iteration of this transformation with the trivial operations of permuting the spatial coordinates x, y, z and of changing the signs of any of the coordinates t, x, y, z, all integral Lorentz transformations (including reflections) are obtained.

This procedure may simplify slightly some of the proofs in this paper. For example, to show that primitive integral null vectors are equivalent, take such a vector (t, x, y, z) and by changing signs make certain that t, x, y, z are all positive or zero. Then so long as $t > 1$ at least two of x, y, z must be non-zero since (t, x, y, z) is assumed primitive. Hence we have

$$t = (x^2 + y^2 + z^2)^{\frac{1}{2}} < x + y + z \leqslant \{3(x^2 + y^2 + z^2)\}^{\frac{1}{2}} < 2t.$$

It follows that $-t < t - x - y - z < 0$. Thus performing (B) and changing signs again, we obtain an integral null vector whose t-component has been decreased. Repeating this process it is clear that we must finally arrive at one of the forms $(1, 1, 0, 0)$, $(1, 0, 1, 0)$, or $(1, 0, 0, 1)$. Permuting the spatial coordinates we can reduce the given primitive integral null vector to the standard form

$$(1, 1, 0, 0). \tag{C}$$

This establishes the theorem of Sec. 7.

University of Toronto

NOTE UPON THE GENERALIZED CAYLEYAN OPERATOR

H. W. TURNBULL

1. The following note which deals with the effect of a certain determinantal operator when it acts upon a product of determinants was suggested by the original proof which Dr. Alfred Young gave of the property

$$(NP)^2 = \theta NP$$

subsisting between the positive P and the negative N substitutional operators, θ being a positive integer[1]. This result which establishes the idempotency of the expression $\theta^{-1}NP$ within an appropriate algebra is fundamental in the Quantitative Substitutional Analysis that Young developed. The present note, which is couched in the language of determinants, proves a result which is equivalent to Young's alternative statement $(PN)^2 = \theta PN$.

These operators P and N take their rise in the theory of groups. In fact let

$$p = p_1 + p_2 + \ldots + p_h$$

be a partition of a positive integer p into h non-zero parts which are arranged in descending order: that is

$$p_1 \geqslant p_2 \geqslant \ldots \geqslant p_h.$$

Let p distinct elements be arranged in the following fashion

$$\begin{array}{l} u_1\ u_2 \ldots\ldots\ldots\ldots u_{p_1} \\ v_1\ v_2 \ldots\ldots\ldots v_{p_2} \\ \ldots\ldots\ldots\ldots \\ w_1\ w_2 \ldots w_{p_h} \end{array}$$

so as to form an array of h rows and p_1 columns, each row being filled consecutively from the left and starting at the first column, while each column is filled consecutively downwards and starts at the first row. No row can exceed in length any row which lies above it, and no column can exceed any column which is upon its left. If $p_1 = p_h$ the array is rectangular: but usually $p_1 > p_h$ and the array has a zigzag boundary upon its right. This array is called a tableau.

Let $f(u_1, \ldots, w_{p_h})$ be a function of these p elements, treated as p arguments of the function, and let $p!$ expressions be formed by interchanging the arguments in every possible way. Usually these expressions will be distinct, as for instance the $2!$ expressions $f(x, y)$ and $f(y, x)$ differ, unless f happens to be symmetric in these two arguments. Let δ_i denote the operation of producing the i^{th} of these expressions, namely

$$\delta_i f(u_1, \ldots, w_{p_h}) = f(u'_1, \ldots, u'_p)$$

where $u'_1, \ldots$ denote the corresponding arrangement of the p arguments

Received February 23, 1948.

[1] [5] p. 366.

$u_1, \ldots w_{p_h}$. There are therefore $p!$ such operations δ_i and they characterize the symmetric group of order $p!$ Let those $p_1!$ distinct operations be performed which permute the elements belonging to the first row only of the tableau. Let the sum of the resulting functions be regarded as the effect of a resultant operation, P_1 say, acting upon the original function: namely

$$P_1f = \Sigma\delta_i f(u_1, \ldots, u_{p_1}, v_1, \ldots) = \Sigma f(u'_1, \ldots, u'_{p_1}, v_1, \ldots)$$

where the first p_1 arguments only are to be permuted, while the remainder are unchanged. This summation has $p_1!$ terms.

Let a corresponding operation be performed for the i^{th} row of the tableau. By taking the rows successively in turn we thus obtain h such operations P_i. Since each of these operations affects a distinct set of arguments, the h operations are independent of one another. We can therefore combine them in any order and form a further resultant operation

$$P = P_1P_2 \ldots P_h = P_2P_1 \ldots P_h = \ldots,$$

which consists of $p_1!\, p_2! \ldots p_h!$ terms, obtained by all the possible different permutations of the elements, each within its own row of the tableau. Since these terms are added together, this P is called the *positive* symmetric group associated with the tableau.

In contrast to this a new operator N_1 is defined, with reference to the first column of the tableau, and consisting of $h!$ terms caused by the complete set of permutations among the elements of this column: only in this case each term that belongs to an odd permutation is accompanied by a negative sign, and otherwise by a positive sign: namely

$$N_1f = \Sigma(-)_j\delta_j f(u_1, \ldots)$$

where the summation has $h!$ terms, δ_j denotes the typical permutation of $u_1, v_1, \ldots w_1$, and $(-)_j$ denotes a positive or negative sign according as the corresponding permutation is even or odd. Let p_1 such operations be defined, one for each column, and combined as before into a resultant operation

$$N = N_1N_2 \ldots N_{p_1} = N_2N_1 \ldots N_{p_1} = \ldots$$

which consists of $q_1!\,q_2! \ldots$ terms, where q_j denotes the number of elements in the j^{th} column $(j = 1, 2, \ldots, p_1)$. This N is called the *negative* symmetric group associated with the tableau. If a row or a column possesses a single element only, the corresponding factor of P or N may be omitted as it has the effect of the factor unity in the whole product.

When a further operator is made by using P and N in succession the products PN and NP usually differ. They do however satisfy the same quadratic relation $X^2 = \theta X$, where $X = PN$ or NP, as already mentioned. One more preliminary remark should be made, before turning to the application of this theory of Young's Substitutional Operators: namely, that the expression $f(u_1, \ldots)$ upon which the operator takes effect may be construed in a most general sense, provided only that each particular arrangement of the p elements $u_1, \ldots$ defines the expression and that they make sense when they are permuted. For instance f might be a determinant, and the u_i might denote suffixes which indicate the columns of the determinant.

The connexion between the abstract analysis and the determinantal theory is as follows. The $n \times n$ determinant $\Sigma \pm x_{11}x_{22} \ldots x_{nn}$ may be written $N_1\phi$, where ϕ is $x_{11}x_{22} \ldots x_{nn}$ and N_1 is the operator which permutes the n second suffixes of the x_{ij} in all possible ways and sums the results accompanied by a negative sign for each interchange of a pair of suffixes. A product ϕ of v determinants may consequently be written $N\phi = N_1N_2 \ldots N_v\phi$, where N_j is the operator which generates the j^{th} determinant from its leading term. If the determinants which compose the product $N\phi$ are not all of the same order, the factors are to be arranged in a descending order. The operation P_1 is then that which generates a sum of $v!$ such terms $N\phi$ by permuting the first columns, one from each of the r determinants, in all their different ways and adding together the results: P_2 likewise permutes all the second columns; and so on until all the columns are so treated. Then $P = P_1P_2 \ldots$, and $PN\phi$ is the final expression. This positive substitutional operation is reflected, in what follows, by taking a single product of determinants and making all *first* columns that occur the *same*; and so on. Except for a factor $v!$ the two expressions are substitutionally equivalent. Again, instead of taking a product ϕ of determinants whose orders may differ, all the factors have been brought up to the same order $n \times n$, by the introduction of arbitrary constant borders, in distinction from which those elements x_{ij} that undergo permutation (or, equivalently, differentiation) are called the variables. Young's formula is implicit in (13) below.

2. Let $x_1, x_2, \ldots, x_n$ denote n sets of n independent variables such that x_i denotes the i^{th} set $\{x_{i1}, x_{i2}, \ldots, x_{in}\}$ when it is arranged in a column. Let $\Delta = (x_1x_2 \ldots x_n)$ denote the $n \times n$ determinant of these n columns in this order, so that Δ is a function of n^2 independent variables x_{ij}. Let $\Omega = (\partial/\partial x_1 \ldots \partial/\partial x_n)$ denote the corresponding determinant when each element x_{ij} is replaced, in its own position, by the corresponding differential operator $\partial/\partial x_{ij}$: thus $\partial/\partial x_i$ denotes the column of the operators which correspond to the column x_i.

Let
$$\sum_{i=1}^{n} \alpha_i \partial/\partial x_{ni} = \left(\alpha \,\middle|\, \frac{\partial}{\partial x_n}\right)$$
denote the polar operator which substitutes a set of n arbitrary constants α_i for the set of variables x_n. Since Δ is a linear form in the n components of x_n it follows at once by differentiation that
$$(\alpha|\partial/\partial x_n)\,\Delta = (x_1x_2 \ldots x_{n-1}\alpha)$$
which we abbreviate to $(X_{n-1}\alpha)$. More generally, and by further such polarizations of the x_i, let r of these sets, say the last r, be replaced by r columns of arbitrary constants, namely

(1) $$\Delta_r = (x_1x_2 \ldots x_{n-r}\beta_1\beta_2 \ldots \beta_r)$$

which we write as $A_Z\Delta = \Delta_r = (X_{n-r}A_r)$, where A_Z denotes the operator which substitutes the block A_r of r columns for Z the block of the last r columns of Δ. (The above single column operator is therefore written as α_z, with $z = x_n$.)

If this is done for the first $n - 1$ values of r we obtain altogether n different

determinants, each of which involves the first column x_1 of the variables, all but one involve the second column x_2, and so on, until Δ alone involves the last column x_n only. Let a power product

$$(2) \qquad \phi = \Delta^{p_0}\Delta_1^{p_1}\dots\Delta_{n-1}^{p_{n-1}} = \prod_{r=0}^{n-1}(X_{n-r}A_r)^{p_r}$$

of these determinants be constructed, where the exponents p_r are zero or positive integers, and where all the blocks of constants A_r are arbitrary.

For example $\phi = (xyz)^p(xy\alpha)^q(x\beta\gamma)^r$ is such a product of three rowed determinants.

It is well known, and indeed it is a fundamental result in the theory of projective invariants, that the effect of the Cayleyan[2] operator $\Omega = |\partial/\partial x_{ji}|$, already mentioned, acting upon a perfect p^{th} power of Δ, is analogous to ordinary differentiation with regard to Δ and yields the identity

$$(3) \qquad \Omega\Delta^p = p(p+1)\dots(p+n-1)\Delta^{p-1}.$$

The object of the present note is to extend this property to the more general power product ϕ, and to shew that

$$(4) \qquad \Omega\phi = p_0(p_0+p_1+1)\dots(p_0+p_1+\dots+p_{n-1}+n-1)\phi_1,$$

where $\phi_1\Delta = \phi$, that is to reduce the index p_0 by unity while leaving the remaining indices unchanged. Naturally if $p_0 = 0$, $\Omega\phi$ vanishes.

To prove this we shall first establish a more general theorem. In fact let a set of positive integers λ_r be introduced where

$$\lambda_r = p_0(p_0+p_1+1)\dots(p_0+p_1+\dots+p_{r-1}+r-1),$$

with $r = 1, 2, \dots, n$. From an $n \times n$ determinant of arbitrary constants let the last r columns be chosen and called B. Furthermore let

$$(5) \qquad B_Z = (b_1b_2\dots b_r \mid \partial/\partial z_1\,\partial/\partial z_2\dots\partial/\partial z_r)$$

denote the bideterminantal (or compound inner product) operator obtained by combining the r columns of B with the last r columns of Ω. Here for convenience the $(n-r+1)^{\text{th}}$ set x has been renamed z_1, and so on until the last x_n is the same as z_r. With this understanding the following result holds:

(6) THEOREM. $B_Z\phi = \lambda_r(X_{n-r}B)\phi_1$.

Proof. We proceed by induction upon r. For if $r = 1$, and b denotes a single column and z denotes x_n, then, by differentiation,

$$b_z\Delta^{p_0} = p_0\Delta^{p_0-1}\,b_z\Delta.$$

But since $\Delta = (X_{n-1}z)$, $\quad b_z\Delta = (X_{n-1}b)$.

Hence $\qquad b_z\Delta^{p_0} = p_0(X_{n-1}b)\,\Delta^{p_0-1}$,

that is $b_z\phi = \lambda_1(X_{n-1}b)\phi_1$ since z is absent from all the remaining factors belonging to ϕ: which proves the result when $r = 1$. By assuming it true for r we shall prove it true for $r+1$. To do this, write

$$X_{n-r} = Xy, \quad y = x_{n-r},$$

so that y denotes the last of the $n-r$ columns x, and X denotes all the earlier columns. The original set of n columns is now exhibited by

$$\Delta = (XyZ).$$

---

[2] [2].

Let $\Delta_0 = (XyB)$. The assumed identity is therefore

$$(7) \qquad B_Z\phi = \lambda_r\Delta_0\Delta^{p_0-1}\Delta_1^{p_1}\ldots\Delta_r^{p_r}N$$

where N denotes all those factors into which the column y does not enter, since in (2) y does not enter Δ_s whenever $s > r$. Now each of the $r+2$ (unrepeated and repeated) factors $\Delta_0, \Delta, \ldots, \Delta_r$ is of the form

$$(XyT)$$

where the block of r columns T differs but X and y are always present in each factor. Operate with c_y, that is $\Sigma c_i\,\partial/\partial y_i$, upon both sides of the equation (7). On the right-hand side we obtain a sum of $r+2$ terms, one for each different Δ. Thus, the affected parts in the various terms are

$$c_y\Delta_0 = (XcB), \quad c_y(XyT)^{p_s} = p_s(XcT)(XyT)^{p_s-1}.$$

Now perform[3] the determinantal permutation $\{c, B\}'$ which consists of $r+1$ terms interchanging c with each of the r columns of B in turn, accompanied by a change of sign, and adding the term (the *static* term, let us say) in which c remains unmoved. The result of this upon the left member $c_yB_Z\phi$ of our equation produces the corresponding operator of order $r+1$, namely

$$\{c, B\}'\,c_yB_Z = \left(cB\,\middle|\,\frac{\partial}{\partial y}\frac{\partial}{\partial Z}\right) = \left(cb_1b_2\ldots b_r\,\middle|\,\frac{\partial}{\partial y}\frac{\partial}{\partial z_1}\frac{\partial}{\partial z_2}\cdots\frac{\partial}{\partial z_r}\right),$$

as is seen at once on expanding this last determinantal expression by its first column.

On the right there are $r+2$ terms, as already seen. In the first term Δ_0 has been altered to (XcB), and in any other term a single Δ_s, say, has been altered to (XcT) multiplied by p_s. The effect of the new operation $\{c, B\}'$ on the first term produces

$$(r+1)\,(XcB)$$

from $C_y\Delta_0$, merely by deranging the $r+1$ columns of cB within this determinant. In each of the other r terms the new operation convolves the columns c, B which occur entirely within Δ_s and Δ_0 respectively. But by the fundamental identity[4]

$$(c, B)'(XyB)(XcT) = (XcB)(XyT)$$

that is, the operation interchanges the c, wherever it occurs with the y, which occurs in Δ_0 the first determinantal factor. This restores the full exponent p_s to Δ_s for $s = 1, 2, \ldots, r$, and in the case of Δ itself restores $p_0 - 1$ which had dropped to $p_0 - 2$ through the operation c_y. Gathering these results together we infer that

$$(8) \quad \left(cB\,\middle|\,\frac{\partial}{\partial y}\frac{\partial}{\partial Z}\right)\phi = \lambda_r[(r+1) + (p_0 - 1) + p_1 + p_2 + \ldots + p_r](XcB)\phi_1$$
$$= \lambda_{r+1}(XcB)\,\phi_1,$$

which is of the same form as the assumed identity but with $r+1$ replacing r. Since the identity is true when $r = 1$ this proves it by induction for $r = 1, 2, \ldots, n$. In the last stage when $r = n-1$ in (8) the operator factorizes into

[3] [4] p. 27. [4] [4] p. 44.

$(cB)\Omega$, and all the columns of X have disappeared. On taking the arbitrary $n \times n$ determinant (cB) to be the unit determinant $|\delta_{ij}|$ the original identity

$$\Omega\phi = \lambda_n\phi_1$$

emerges.

COROLLARY 1. The same identity is true if each A_r that occurs is replaced by p_r arbitrary blocks A'_r, A''_r, etc. This follows since, in the above proof, no use is made of the value of T, but only of its extent.

3. On writing

$$
\begin{aligned}
&p_0 + p_1 + \ldots + p_{n-1} = q_1 \\
&p_0 + p_1 + \ldots + p_{n-2} \quad = q_2 \\
(9) \qquad &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&p_0 \qquad\qquad\qquad\qquad\quad = q_n
\end{aligned}
$$

where $q_i = p_0 + p_1 + \ldots + p_{n-i}$, we obtain the numbers of times which each x_i occurs in the product ϕ, x_i occurring exactly q_i times, for $i = 1, 2, \ldots, n$. In particular $z = x_n$, and the last column of Δ, appears q_n times. Accordingly, if we operate q_n times in succession with Ω and apply the theorem, we obtain

$$(10) \qquad \Omega^{q_n}\phi = \mu_0\Delta_1{}^{p_1}\Delta_2{}^{p_2}\ldots\Delta_{n-1}^{p_{n-1}} = \mu\psi_1, \text{ say,}$$

where μ_0 is a product of positive integers λ_r, and from which the last column x_n has disappeared. From (4) we obtain

$$
\begin{aligned}
\mu_0 &= p_0!\,\frac{(p_0 + p_1 + 1)!}{(p_1 + 1)!}\cdots\frac{(p_0 + p_1 + \ldots + p_{n-1} + n - 1)!}{(p_1 + \ldots + p_{n-1} + n - 1)!} \\
(11) \qquad &= q_n!\,\frac{(q_{n-1} + 1)!}{(p_1 + 1)!}\cdots\cdots\cdots\frac{(q_1 + n - 1)!}{(q_1 + n - 1 - p_0)!}.
\end{aligned}
$$

Now let Ω_1 denote the $(n - 1)$-fold column of operators (each component being a determinant of order $n - 1$):

$$\{\,|\partial/\partial x_1\,\partial/\partial x_2\ldots.\partial/\partial x_{n-1}|\,\}$$

and let $C_X = (C\,|\,\Omega_1)$ be such as the operator (5) but with $r = n - 1$. Then, by the theorem, $C_X\psi_1$ reduces the exponent p_1 of Δ_1 by unity, and C_X applied p_1 times replaces the X in this factor by C, and introduces a positive integral factor

$$\mu_1 = p_1!\,\frac{(p_1 + p_2 + 1)!}{(p_2 + 1)!}\cdots\frac{(p_1 + \ldots + p_{n-1} + n - 2)!}{(p_2 + \ldots + p_{n-1} + n - 2)!}.$$

Now write C_{n-1} for this block C. We may proceed in this way with further operators $(C_{n-r}\,|\,\Omega_r)$ of this type, where $r = n - 2$, $n - 3$ and so on, in succession: for the theorem is directly applicable at each such stage, and replaces all the X_{n-r} in $\Delta_r{}^{p_r}$ by an equal number of C_{n-r} while attaching a further positive integral factor μ_r. If preferred all the C_{n-r} which are p_r in number can be distinct, for they are arbitrary. The whole operation can now be written

$$\Omega\prod_{s=1}^{n-1}(C_{n-s}\,|\,\Omega_s)$$

and, since it is composed entirely of differential operators $\partial/\partial x_{ji}$ and constants, the order of its factors is immaterial. On retaining the original $n \times n$ constant

determinant (now called $|C_n|$) along with Ω we may drop the factor Ω and let s run from 0 to $n-1$ in the product; for $(C_n|\,\Omega_0)$ factorizes to $|\,C_n|\,\Omega$ since $|\Omega_0| = \Omega$. It is then convenient to express the whole operator in terms of the integers q_i as follows:

$$(12) \qquad \prod_{s=0}^{n-1} (C_{n-s}\,|\,\Omega_s) \equiv (C_{q_1q_2\,\cdots\,q_n}\,|\,\Omega_{q_1q_2\,\cdots\,q_n}) \equiv (C_Q\,|\,\Omega_Q) \equiv C_\Omega\,.$$

Here the capital suffix Q denotes the multiple suffix $q_1q_2\ldots$: and in the latter, which is a set of positive integers written in descending order since their first differences are the p_i which are $\geqslant 0$, it is unnecessary to include any zero suffixes. This Q therefore denotes a partition $\{q_1q_2\ldots\}$ of Σq_i, written in the usual way. Reference to (2) shews that the operator effects the substitution of the C_Q's for the X's as follows:

$$(13) \qquad \prod_{s=0}^{n-1} (C_{n-s}\,|\,\Omega_s)\phi = \theta_Q \prod_{r=0}^{n-1} (C_{n-r}A)_r{}^{p_r}\,,$$

where $\theta_Q = \mu_0\mu_1\ldots\mu_r\ldots\mu_{n-1}$, a numerical constant which is a positive integer. The more general case when all p_r of the C_{n-r} are distinct, for each value of r, can be written down without serious difficulty (only it is rather prolix !). It has the same numerical factor θ_Q.

Two further corollaries follow at once:

Corollary 2. Take all the A_r which occur in the product ϕ to be non-zero portions of the unit matrix $[\delta_{ij}]$, so that $(X_{n-r}A_r)$ is then an $(n-r)$ rowed minor of the determinant $|x_{ij}|$. Thus ϕ is a power product of such minors of all orders (every minor of a lower order being a minor within the columns but not necessarily the rows occupied by a minor of a higher order, owing to the original condition imposed upon the columns x_i). Take each C_{n-r} to be the complementary portion of the unit matrix so that $(C_{n-r}A_r) = 1$. Then the corresponding operator C_Ω reduces ϕ to the positive integer θ_Q.

Corollary 3. Replace each C_{n-s} by the corresponding matrix X_{n-s}. Then $(X_{n-s}|\,\Omega_s)$ is the well-known *Capelli operator*.[5] The generalized operator X_Ω will produce two sorts of terms when it operates on any function of the x_{ij}—(i) intrinsic terms due to differentiating those parts X_t of the operator which stand in factors to the right of the partial operator $\partial/\partial x_{ij}$, and (ii) extrinsic terms due to direct operation on the operand. Since the right-hand side member of (13) reverts to ϕ itself on substituting the X for the C, it follows that

$$\text{extr}\, X_\Omega\phi = \theta_Q\phi$$

where the notation indicates the extrinsic terms only.

What happens to the intrinsic terms? Is there a result comparable in beauty to the original formula of Capelli? This formula expresses the operator

$$\sum_I (x_1x_2\ldots x_s)_I(\partial/\partial x_1\, \partial/\partial x_2 \ldots \partial/\partial x_s)_I$$

(for $I = i_1i_2\ldots i_s$, any set of s different integers $1, 2, \ldots, n$) as a determinant

$$|\,(x_i|\,\partial/\partial x_j) + (n-i)\delta_{ij}\,|\,, \qquad i, j = 1, 2, \ldots n,$$

[5] [1].

where the first $n-1$ integers appear in descending order, finishing with zero, as additions to the elements upon the leading diagonal. These additions are caused by the intrinsic terms, and the expansion of the whole determinant must be taken in the strict order of its columns.[6]

As an example of the complete operator acting upon

$$\phi = (xyz)(xy\alpha)^2(x\beta\gamma)$$

take $C_\Omega \equiv (\partial/\partial x\, \partial/\partial y\, \partial/\partial z)\,.\,(\delta\epsilon \mid \partial/\partial x\, \partial/\partial y)^2(\zeta \mid \partial/\partial x)$ where δ, ϵ, ζ are arbitrary columns, and all the columns consist of three elements each. Then Q, as in (12), denotes the suffix row 4, 3, 1 which indicate the numbers of appearances of x, y, z respectively in ϕ. Then

$$C_\Omega = \theta_{431}(\delta\epsilon\alpha)^2(\zeta\beta\gamma) = 576(\delta\epsilon\alpha)^2(\zeta\beta\gamma).$$

Again, if $\alpha\beta\gamma\delta\epsilon\zeta$ denote the columns of the unit matrix, the result is zero unless $\delta\epsilon\alpha$ include the three different columns, as well as $\zeta\beta\gamma$. For instance

$$(\partial/\partial x\, \partial/\partial y\, \partial/\partial z)(\partial/\partial x\, \partial/\partial y)^2_{13}(\partial/\partial x)_1\, \phi = 576$$

when $\phi = (xyz)(xy)^2_{13}x_1$.

The numerical coefficient

$$\theta_Q \equiv \theta_{q_1q_2\cdots q_n}$$

may be found from the above product $\mu_0\mu_1 \ldots \mu_{n-1}$ where

$$\mu_i = p_i!\,\frac{(p_i+p_{i+1}+1)!}{(p_{i+1}+1)!}\cdots\frac{(p_i+\ldots+p_{n-1}+n-i-1)!}{(p_{i+1}+\ldots+p_{n-1}+n-i-1)!},$$

for these actual values of the μ_i follow directly by repeated use of the identity (6). On substituting for the μ_i in terms of the q_i we obtain[7]

$$(14)\qquad \theta_Q = \Pi\mu_i = \frac{\Pi\,(q_r+r-1)!}{\prod\limits_{r<s}(q_r-q_s-r+s)},\qquad \left.\begin{matrix} r\\ s\end{matrix}\right\} = 1, 2, \ldots, n\,.$$

This is a positive integer, since each μ_i is, and it is the well-known cofactor of the number f_Q for $n!$, namely

$$\theta_Q f_Q = n!$$

where f_Q is the group characteristic $\chi_0{}^Q$, or the Q^{th} component of the character χ_0, which was given by Frobenius for the symmetric group of order $N = \Sigma q_i$.

This number θ_Q can also be defined[8] by the determinant

$$\Delta_Q = \left| d_{ij} \right| = \left| \frac{1}{(q_i-i+j)!} \right| = \frac{1}{\theta_Q},$$

where $d_{ij} = 0$ whenever $q_i < i-j$ and $d_{ij} = 1$ whenever $q_i = i-j$. The number of rows and columns in the determinant is taken to be the number of non-zero suffixes in the set $Q = q_1q_2 \ldots q_n$ ($q_1 \geqslant q_2 \geqslant$ etc.).

---

[6] Cf. [1], and [4] p. 117.

[7] Cf. [5] p. 366.

[8] [4] p. 359. A misprint is here corrected from j to $i-j$.

For example

$$\frac{1}{\theta_{431}} = \begin{vmatrix} \frac{1}{4!} & \frac{1}{5!} & \frac{1}{6!} \\ \frac{1}{2!} & \frac{1}{3!} & \frac{1}{4!} \\ 0 & 1 & \frac{1}{1!} \end{vmatrix} = \frac{1}{576}.$$

The proof that the determinantal and the product formulae for θ_Q are equivalent follows at once on evaluating the determinant by Dodgson's method.[9] If by compact minor we mean a minor chosen in any manner from any r consecutive rows and any r consecutive columns of the original determinant, then the method depends upon the systematic condensation of Δ_Q by the use of compact minors. Here for instance, if u_Q denotes $1/\theta_Q$, we have the condensation

$$u_{431} = \begin{vmatrix} u_4 & u_5 & u_6 \\ u_2 & u_3 & u_4 \\ 0 & 1 & u_1 \end{vmatrix}, \quad \begin{vmatrix} u_{43} & u_{54} \\ u_{21} & u_{31} \end{vmatrix}, \quad \begin{vmatrix} u_{431} \end{vmatrix} = u_{431}.$$

In this and all such sequences of condensing determinants those elements which stand *within* the whole outer border of elements are called pivotal elements (u_3 alone is such a pivot in this example). If v is such a pivot and V is the 3×3 minor determinant of which v is the central element then V appears as an element in the next but two member of the sequence. If v happens to be zero, then by definition of Δ_Q, the three consecutive elements which stand in the row immediately below that of v,

$$\begin{matrix} & 0 & & \ldots \\ 0 & 0 & 0 & \ldots \end{matrix}$$

symmetrically, must also be zero. Thus V also vanishes. When $v \neq 0$ the usual pivotal process is available (for example $u_{43}u_{31} - u_{54}u_{21} = u_{431}u_3$, where u_3 can be cancelled since it does not vanish). In either case the process is definite, and leads to the required result.

References

[1] A. Capelli, *Math. Ann.*, vol. 29 (1887), 331-338.

[2] A. Cayley, *Collected Works*, vol. 1 (1845), 80-94, 95-112.

[3] J. H. Grace and A. Young, *The Algebra of Invariants* (Cambridge, 1903), 259.

[4] H. W. Turnbull, *Theory of Determinants, Matrices and Invariants* (Glasgow, 1928; 2nd ed. 1945).

[5] A. Young, "Quantitative Substitutional Analysis," *Proc. London Math. Soc.* (1) vol. 34 (1902), 361-397, in particular p. 364.

The University
St. Andrews, Scotland

[9] [4] p. 340.

ELEMENTARY ALGEBRAIC TREATMENT OF THE QUANTUM MECHANICAL SYMMETRY PROBLEM

HERMANN WEYL

1. Stating the problem

A function $\eta(i_1, \ldots, i_f)$ of f quantities i, varying over the finite range $i = 1,2, \ldots, n$, is usually called an n-dimensional tensor of rank f. Any permutation p: $1 \to 1', \ldots, f \to f'$ changes this tensor into a tensor $\mathbf{p}\eta$ according to the equation $\mathbf{p}\eta(i_1, \ldots, i_f) = \eta(i_{1'}, \ldots, i_{f'})$. Thus the permutation p appears as a linear operator $\mathbf{p}$ in the n-dimensional space $\Sigma = \Sigma_{n,f}$ of all n-dimensional tensors of rank f. η is symmetric if $\mathbf{p}\eta = \eta$ for all permutations p, it is antisymmetric if $\mathbf{p}\eta = \delta_p . \eta$ where $\delta_p = +1$ for the even and -1 for the odd permutations. Let a linear transformation A in Σ,

(1.1) $\quad \eta' = A\eta, \quad \eta'(i_1 \ldots i_f) = \sum_k a(i_1 \ldots i_f\,; k_1 \ldots k_f) \cdot \eta(k_1 \ldots k_f),$

be called symmetric [1] if

$$a(i_{1'} \ldots i_{f'}\,; k_{1'} \ldots k_{f'}) = a(i_1 \ldots i_f\,; k_1 \ldots k_f)$$

for all permutations p. A is symmetric if and only if it commutes with all the permutation operators $\mathbf{p}$. The symmetric transformations A form an algebra $\mathfrak{A}$. The general symmetry problem posed by the quantum theory of an aggregate of f equal physical entities is this:

(I) *to decompose the tensor space* Σ *as far as possible into subspaces* Π *that are invariant with respect to all symmetric transformations* A.

An epistemological principle basic for all theoretical science, that of projecting the actual upon the background of the possible, is here followed by asking what happens under any *possible* Schrödinger law of dynamics $\frac{h}{i}\frac{d\eta}{dt} = A\eta$, before taking up the specific law involving the *actual* energy operator $A = H$. We have here ignored the further condition which physics imposes on all energy operators A, to wit their Hermitean nature,

$$a(k_1 \ldots k_f\,; i_1 \ldots i_f) = \bar{a}(i_1 \ldots i_f\,; k_1 \ldots k_f).$$

Essential for the theory of eigenvalues (terms) and eigenfunctions, this condition is irrelevant for our purposes. For what is invariant under all Hermitean symmetric transformations stays so even when the Hermitean restriction is lifted. As algebraists we are glad to get rid of it. For we propose to carry our investigation through in any number field in which the equation $f!a = 0$ for a number a implies $a = 0$ (field of characteristic 0 or of a prime characteristic dividing none of the natural numbers $1, 2, \ldots, f$).

Received February 28, 1948.

[1] We shall adhere to this terminology and not use the word symmetric in the sense presently to be mentioned under the name Hermitean.

It is no wonder that the complete solution of the above symmetry problem depends on the theory of representations of the symmetric group of all permutations and Young's symmetry operators.[2]

Let Σ^+, Σ^- denote the linear manifolds of all symmetric or antisymmetric tensors respectively. Nature has most wisely put a stop to the breaking-up of Σ into isolated compartments Π by letting but one of them, the invariant subspace Σ^-, come into existence. Such at least is the case if the f entities of which the aggregate is composed are electrons (Pauli's exclusion principle). Thereby the symmetry problem (I) loses its significance for physics. Part of it, however, is restored, if the existence of the spin of the electron is taken into account but its dynamical influence disregarded—a procedure which is at least approximately permissible. The situation is then as follows. The argument i is replaced by a pair $(i\rho)$ with the range $i = 1, \ldots, n$ for the "positional" variable i and the range $\rho = 1, \ldots, \nu$ for the "spin" variable ρ. (Actually $\nu = 2$ while the positional variable varies over the continuum of all possible positions in the physical three-dimensional space.) Set $N = n\nu$. The possible wave states of the aggregate of f electrons are described by the antisymmetric N-dimensional tensors $\psi(i_1\rho_1, \ldots, i_f\rho_f)$ of rank f, forming the space $\Sigma^-{}_{N,f} = \Omega$. Moreover we envisage the space $\Sigma = \Sigma_{n,f}$ of all n-dimensional tensors $\eta(i_1 \ldots i_f)$ of rank f, and the space $\mathrm{P} = \Sigma_{\nu,f}$ of all ν-dimensional tensors $\phi(\rho_1, \ldots, \rho_f)$ of rank f. Any symmetric transformation A in Σ,

$$\eta'(i_1 \ldots i_f) = \textstyle\sum_k a(i_1 \ldots i_f\,;\,k_1 \ldots k_f) \cdot \eta(k_1 \ldots k_f)$$

induces a transformation A^* in Ω,

$$\psi'(i_1\rho_1, \ldots, i_f\rho_f) = \textstyle\sum_k a(i_1 \ldots i_f\,;\,k_1 \ldots k_f) \cdot \psi(k_1\rho_1, \ldots, k_f\rho_f).$$

The central problem is

(II) *to decompose Ω as far as possible into subspaces that are invariant under the transformations A^* thus induced in Ω by all symmetric transformations A in Σ.*

These A^* form an algebra $\mathfrak{A}^*$. It is also true that any symmetric transformation B in P,

$$\phi'(\rho_1 \ldots \rho_f) = \textstyle\sum_\sigma b(\rho_1 \ldots \rho_f\,;\,\sigma_1 \ldots \sigma_f) \cdot \phi(\sigma_1 \ldots \sigma_f), \tag{1.3}$$

induces a corresponding transformation B^* in Ω,

$$\psi'(i_1\rho_1, \ldots, i_f\rho_f) = \textstyle\sum_\sigma b(\rho_1 \ldots \rho_f\,;\,\sigma_1 \ldots \sigma_f) \cdot \psi(i_1\sigma_1, \ldots, i_f\sigma_f). \tag{1.4}$$

The B^* form an algebra $\mathfrak{B}^*$. Every A^* of $\mathfrak{A}^*$ commutes with every B^* of $\mathfrak{B}^*$.

Not only the problem (I), but also this new symmetry problem (II) may be solved by means of Young's symmetry operators; cf. GQ, chap. v, § 12. However, as shall be discussed here in detail, a more elementary approach is available for the physically important case $\nu = 2$. Indeed the decomposition of the spin tensor space $\mathrm{P} = \Sigma_{2,f}$ into irreducible invariant subspaces under the algebra $\mathfrak{B}$ of all its symmetric transformations B is readily derived from

[2] Cf. H. Weyl, *Gruppentheorie und Quantenmechanik* (2nd ed. Leipzig, 1931) [quoted as GQ], chap. V, §§ 1-7 and 13-14.

the classical Clebsch-Gordan expansion. From the algebra $\mathfrak{B}$ in P we may pass to its representation $\mathfrak{B}^*$ in Ω. Because of the commutability of the elements A^* and B^* of $\mathfrak{A}^*$ and $\mathfrak{B}^*$, decomposition of the generic matrix of $\mathfrak{B}^*$ entails a "dual" decomposition for $\mathfrak{A}^*$. The deeper lying fact that vice versa any linear transformation in Ω that commutes with all $B^* \epsilon \mathfrak{B}^*$ lies in $\mathfrak{A}^*$ is needed in order to show that the latter decomposition is also one into irreducible parts.

All linear transformations (matrices) in a g-dimensional vector space Ξ form an algebra $\mathfrak{M}_g$ of order g^2, the complete matric algebra of degree g. Throughout our investigation irreducibility for matric algebras will be sharpened to completeness. Decomposition of a matrix C into two matrices $C_1|C_2$ is defined by the equation

$$C = \left\| \begin{matrix} C_1 & 0 \\ 0 & C_2 \end{matrix} \right\|$$

$1 \circ C$, $2 \circ C$, $3 \circ C$, ... are the abbreviations for C, $C|C$, $C|C|C$, ..., and $\mathbf{S}$ is the summation sign for the addition $|$ of matrices. Let $\mathfrak{C}$ be a matric algebra of order m in a g-dimensional vector space Ξ. Suppose that, relative to a suitably chosen coordinate system for Ξ, the generic matrix C of $\mathfrak{C}$ decomposes into $m_1 \circ C_1 | m_2 \circ C_2 | \ \ldots$, the matrix C_r of degree g_r occurring with the multiplicity $m_r > 0$, $g = m_1 g_1 + m_2 g_2 + \ldots$. The $g_1^2 + g_2^2 + \ldots$ coefficients of the matrices $C_1, C_2, \ldots, C_h$ are linear forms of the m parameters of $\mathfrak{C}$. We speak of *complete decomposition* if these coefficients are all linearly independent and thus $m = g_1^2 + g_2^2 + \ldots$. For $\nu = 2$ we shall prove the following

Main Theorem. *Relative to a suitably chosen coordinate system for the space Ω, the generic matrix A^* of $\mathfrak{A}^*$ suffers complete decomposition*

$$A^* = \mathbf{S}(v + 1) \circ A^*_u ; \tag{1.5}$$

u and v are two non-negative integers related by the equation $2u + v = f$. The part A^*_u of "valence defect" u and the corresponding "valence" v occurs with the multiplicity $v + 1$. Set $d = n - f$, $\bar{u} = d + u$. The degree g^*_u of the matrix A^*_u is given by the formula

$$g^*_u = \binom{n}{u}\binom{n}{\bar{u}} \frac{(n+1)(n+1-u-\bar{u})}{(n+1-u)(n+1-\bar{u})}, \tag{1.6}$$

$\binom{n}{u}$ denoting the binomial coefficient $\dfrac{n!}{u!(n-u)!}$. Only those u occur in the sum (1.5) for which $u \geq 0$, $\bar{u} \geq 0$, $v = n - (u + \bar{u}) \geq 0$.

Spectroscopically this theorem establishes the existence of non-intercombining term systems corresponding to the various valences v. The terms of valence v are of multiplicity $v + 1$. Only when the actually existing weak interactions between the spins are taken into account, each term of valence v splits into a "multiplet" of $v + 1$ slightly different terms; whereas the weak interaction between position and spin accounts for weak intercombinations between the

several term systems. The significance of the valence v for chemistry is sufficiently indicated by its name.

After some preliminaries in **2** the decomposition (1.5) is derived from the Clebsch-Gordan expansion in **3**. Its completeness will be proved in **4** and **5**.

2. Auxiliary propositions

Schur's lemma for complete instead of irreducible matric algebras is a triviality; nevertheless it may be stated as our

LEMMA 1. Complete decomposition of the generic matrix C of a matric algebra $\mathfrak{C}$, $C = m_1{}^\circ C_1 | m_2{}^\circ C_2 | \ldots | m_h{}^\circ C_h$, implies the same for its commutator algebra $\mathfrak{D}$, $D = g_1{}^\circ D_1 | g_2{}^\circ D_2 | \ldots$. But degree and multiplicity are interchanged: the degree g_r of C_r is the multiplicity with which D_r occurs in the generic matrix D of $\mathfrak{D}$, and the multiplicity m_r of C_r is the degree of D_r.

As one knows, the commutator algebra of a given matric algebra $\mathfrak{C}$ consists of those matrices D that commute with all elements C of $\mathfrak{C}$. As an abstract algebra $\mathfrak{c}$ the completely decomposed matric algebra $\mathfrak{C}$ of Lemma 1 is the direct sum of a number of complete matric algebras; indeed $\mathfrak{c}$ consists of all h-uples $(C_1, \ldots, C_h)$ of arbitrary matrices $C_1, \ldots, C_h$ of the respective degrees $g_1, \ldots, g_h$. We need the following classical proposition, for the simple proof of which I refer the reader to GQ, p. 271, Satz (6.1).

LEMMA 2. Every representation of the direct sum $\mathfrak{c}$ of h complete matric algebras is of the form

$$(C_1, \ldots, C_h) \to m^*{}_1{}^\circ C_1 | \ldots | m^*{}_h{}^\circ C_h.$$

(Here some of the multiplicities $m^*{}_r$ may be zero; this will happen if the representation is not faithful and hence the representing matric algebra $\mathfrak{C}^*$ is of lower order than $\mathfrak{C}$.)

Any *antisymmetric* n-dimensional tensor η of rank f is completely characterized by its components $\eta(\iota_1 \ldots \iota_f)$ with $\iota_1 < \iota_2 < \ldots < \iota_f$, and these are independent. We have

$$\eta(i_1 \ldots i_f) = \delta_i \cdot \eta(\iota_1 \ldots \iota_f)$$

for any permutation $i_1 \ldots i_f$ of $\iota_1 \ldots \iota_f$, $\delta_i = \pm 1$ distinguishing the even from the odd permutations $\begin{pmatrix} \iota_1 \ldots \iota_f \\ i_1 \ldots i_f \end{pmatrix}$, and $\eta(i_1 \ldots i_f) = 0$ if the numbers $i_1 \ldots i_f$ are not all distinct. Hence $\Sigma^- = \Sigma^-{}_{n,f}$ does not exist unless $n \geq f$, and its dimensionality is

$$M^-{}_n(f) = \frac{n!}{f!(n-f)!}.$$

LEMMA 3. Any linear transformation in Σ^- may be written in the form (1.1) where $a(i_1 \ldots i_f ; k_1 \ldots k_f)$ is antisymmetric in the f arguments i, antisymmetric in the f arguments k [and hence symmetric in the f pairs (ik)].

Indeed a linear transformation in Σ^-,

$$\eta'(\iota_1 \ldots \iota_f) = \textstyle\sum_\kappa a(\iota_1 \ldots \iota_f ; \kappa_1 \ldots \kappa_f) \cdot \eta(\kappa_1 \ldots \kappa_f)$$

(with the sum extending over the possible sequences $\kappa_1 < \ldots < \kappa_f$ chosen from the range $1, 2, \ldots, n$) may be written as (1.1) when one puts

$$a(i_1 \ldots i_f ; k_1 \ldots k_f) = \frac{\delta_i \delta_k}{f!} \cdot a(\iota_1 \ldots \iota_f ; \kappa_1 \ldots \kappa_f)$$

for any permutation $i_1 \ldots i_f$ of $\iota_1 \ldots \iota_f$ and any permutation $k_1 \ldots k_f$ of $\kappa_1 \ldots \kappa_f$, and puts $a(i_1 \ldots i_f ; k_1 \ldots k_f) = 0$ in case the numbers $i_1 \ldots i_f$ or $k_1 \ldots k_f$ are not all distinct.

It follows from this lemma that the algebra $\mathfrak{A}$ of symmetric transformations is a complete matric algebra *in the invariant subspace* Σ^- of Σ.

Any *symmetric* tensor η may be completely characterized by its components $\eta(i_1 i_2 \ldots i_f)$ with $i_1 \leq i_2 \leq \ldots \leq i_f$, and these are independent. On changing the labels $i_1 \ldots i_f$ into $i_1 + 0, i_2 + 1, i_3 + 2, \ldots, i_f + (f-1)$ one sees at once that the dimensionality of the space $\Sigma^+ = \Sigma^+_{n,f}$ of symmetric tensors equals

$$M^+_n(f) = \frac{(n + f - 1)!}{f!\ (n - 1)!}.$$

Set $\eta(i_1 \ldots i_f) = \eta_{f_1 \ldots f_n}$ if f_1 of the f arguments $i_1 \ldots i_f$ equal 1, f_2 of them equal 2, $\ldots$, f_n of them equal n. These numbers $\eta_{f_1 \ldots f_n}$ corresponding to the various partitions $f_1 + f_2 + \ldots f_n$ of f can also be used as the independent components of η. A typical symmetric tensor arises from a vector $(x_1, \ldots, x_n)$ by the formula

(2.1) $$\eta(i_1 \ldots i_f) = x_{i_1} \ldots x_{i_f} \text{ or } \eta_{f_1 \ldots f_n} = x_1^{f_1} \cdots x_n^{f_n}.$$

A linear form $l(\eta)$ depending on a variable symmetric tensor η is to be written as

$$l(\eta) = \Sigma l_{f_1 \ldots f_n} \cdot \eta_{f_1 \ldots f_n}$$

with a constant coefficient $l_{f_1 \ldots f_n}$ for each partition $f_1 + \ldots + f_n$ of f. We make the altogether trivial remark that $l(\eta)$ vanishes identically in η provided it vanishes identically in x by dint of the substitution (2.1).

The symmetric transformation $B = \|b(\rho_1 \ldots \rho_f ; \sigma_1 \ldots \sigma_f)\|$ of the algebra $\mathfrak{B}$ may be looked upon as a symmetric ν^2-dimensional tensor $b(\omega_1, \ldots, \omega_f)$ of rank f, if each pair $(\rho\sigma)$ is taken as a single argument ω capable of ν^2 values. Hence the order of the matric algebra $\mathfrak{B}$ in P is $M^+_{\nu^2}(f)$ [and the order of $\mathfrak{A}$ is $M^+_{n^2}(f)$]. The linear transformation $t = \| t_{\rho\sigma} \|$ in the ν-dimensional vector space induces the symmetric transformation $B(t)$,

(2.2) $$b(\rho_1 \ldots \rho_f ; \sigma_1 \ldots \sigma_f) = t_{\rho_1\sigma_1} \cdots t_{\rho_f\sigma_f}$$

in the tensor space P. Considering the ν^2 coefficients $t_{\rho\sigma}$ as indeterminates, we speak of t as the generic element of the linear group ζ and of $t \to B(t)$ as the representation ζ^f of ζ. Equation (2.2) is in complete analogy to (2.1), and the "altogether trivial remark" made above amounts to the following

LEMMA 4. A linear form $l(B)$ depending on an arbitrary element B of $\mathfrak{B}$ vanishes identically if it vanishes identically in the parameters $t_{\rho\sigma}$ for $B = B(t)$.

As a final lemma we write down a simple formula for the case $\nu = 2$, $\nu^2 = 4$:

LEMMA 5.

$$M^{+}{}_4(f) = \frac{(f+1)(f+2)(f+3)}{1\cdot 2\cdot 3} = \Sigma(v+1)^2 \tag{2.3}$$

where the sum extends over the non-negative members v of the sequence $f, f-2, f-4, \ldots$.

Proof. Verify (2.3) for $f = 0, 1$ and the relation

$$M^{+}{}_4(f) - M^{+}{}_4(f-2) = (f+1)^2$$

for all $f \geq 2$.

3. The Clebsch-Gordan expansion and the decomposition of $\mathfrak{A}^*$

In this section we assume $\nu = 2$.

The symmetric 2-dimensional tensors $\phi(\rho_1 \ldots \rho_v)$ $(\rho = 1,2)$ of rank v $(\leq f)$ form a linear manifold $\mathrm{P}^{+}{}_v = \Sigma^{+}{}_{2,v}$ of $v+1$ dimensions. In agreement with a usage established above denote by ϕ_h the component $\phi(\rho_1 \ldots \rho_v)$ in which h of the v arguments ρ have the value 1 and $h - v$ have the value 2 $(h = 0,1,\ldots, v)$. The indeterminate transformation $t = \| t_{\rho\sigma} \|$ $(\rho,\sigma = 1,2)$ in the 2-dimensional vector space induces the transformation

$$\phi'(\rho_1 \ldots \rho_v) = \Sigma_\sigma t_{\rho_1\sigma_1} \ldots t_{\rho_v\sigma_v} \cdot \phi(\sigma_1 \ldots \sigma_v)$$

in $\mathrm{P}^{+}{}_v$, and thus $\mathrm{P}^{+}{}_v$ appears as the representation space of a definite representation Z_v of ζ of degree $v+1$. By multiplying the transformed components ϕ'_h by a fixed power Δ^u $(u = 0, 1, 2, \ldots)$ of the determinant $\Delta = t_{11}\, t_{22} - t_{12}\, t_{21}$ one obtains a representation $\Delta^u \mathrm{Z}_v$ of ζ of the same degree $v + 1$. Envisage the subgroup ζ_0 of ζ, the generic element of which is the substitution

$$\begin{Vmatrix} t_{11}, & t_{12} \\ t_{21}, & t_{22} \end{Vmatrix} = \begin{Vmatrix} \lambda, & 0 \\ 0, & 1 \end{Vmatrix} \tag{3.1}$$

with one indeterminate parameter λ. That substitution multiplies ϕ_h by λ^h according to the representation Z_v, by λ^{u+h} according to the representation $\Delta^u \mathrm{Z}_v$. Hence the coordinates in the representation space Π of $\Delta^u \mathrm{Z}_v$ are so chosen that they are distinguished by a signature ("magnetic quantum number") $w = u + h$. This signature is the exponent of the factor λ^w taken on by the coordinate with the label w under the influence of (3.1) and ranges over the values $w = u, u+1, \ldots, u+v$. [Decomposition of Π into one-dimensional parts invariant with respect to the subgroup ζ_0 of ζ.]

The 2-dimensional tensors $\phi(\rho_1 \ldots \rho_a, \rho_{a+1} \ldots \rho_{a+b})$ of rank $a + b$ $(\leq f)$ which are symmetric in the first a and symmetric in the last b arguments form the substratum of the representation $\mathrm{Z}_a \times \mathrm{Z}_b$ of ζ of degree $(a+1)(b+1)$. The latter breaks up into parts in accordance with the Clebsch-Gordan formula

$$\mathrm{Z}_a \times \mathrm{Z}_b = \mathsf{S}\Delta^u \mathrm{Z}_v, \tag{3.2}$$

the sum extending over all non-negative integers u, v for which $2u + v = a + b$ and $u \leq \min(a, b)$. This follows by induction from the equation

$$\mathrm{Z}_a \times \mathrm{Z}_b = \mathrm{Z}_{a+b} \,|\, \Delta(\mathrm{Z}_{a-1} \times \mathrm{Z}_{b-1}).$$

A simple proof is to be found, for instance, on pp. 115-117 of GQ.

Repeated application of (3.2) leads to a formula of this type:

$$Z_1 \times Z_1 \times \ldots \times Z_1 \ (f \text{ factors}) = \mathfrak{S} g_u \circ \Delta^u Z_v \qquad (2u + v = f).$$

$Z_1 \times \ldots \times Z_1$ is nothing but the representation ζ^f, $t \to B(t)$, of ζ in P, and our formula states that the matrix $B(t)$ breaks up in the manner described by

$$B(t) = \mathfrak{S} g_u \circ B_u(t) \tag{3.3}$$

into partial matrices $B_u(t)$ of degree $v + 1$. Here u, v range over all non-negative integers satisfying the equation $2u + v = f$, and each component $B_u(t)$ occurs with a certain multiplicity $g_u \geq 0$.

If we now make use of Lemma 4, which also states that two linear forms $l(B)$ are identical if they become identical by the substitution $B = B(t)$, we see at once that the generic matrix B of $\mathfrak{B}$ itself breaks up in the same fashion

$$B = \mathfrak{S} g_u \circ B_u. \tag{3.4}$$

Lemma 5 then shows that none of the valences $v = f, f - 2, f - 4, \ldots$ is left out, $g_u > 0$ for $0 \leq u \leq \frac{1}{2}f$, and that all the coefficients of the various matrices $B_{\overline{u}}^{u}$ are independent linear forms of the $M^+{}_4(f)$ parameters $b_{f_1 f_2 f_3 f_4}$ of B. Hence (3.4) is a *complete decomposition*.

$\mathfrak{B}^*$ is a representation of $\mathfrak{B}$, and thus Lemma 2 leads to a similar formula

$$B^* = \mathfrak{S} g^*{}_u \circ B_u \qquad (g^*{}_u \geq 0) \tag{3.5}$$

for the generic matrix B^* of $\mathfrak{B}^*$.

It is not difficult to determine the multiplicities $g^*{}_u$ explicitly. Specialize the element t of ζ by (3.1) in $B = B(t)$ and the corresponding $B^*(t)$. The effect of this specialized $B^*(t)$ upon a tensor component $\psi(i_1\rho_1, \ldots, i_f\rho_f)$ is multiplication by λ^w if w of the f indices $\rho_1, \ldots, \rho_f$ are 1 (and $f - w$ of them equal 2). A complete set of independent components of ψ of that type is obtained by choosing

$$\left.\begin{array}{l} \rho_1 = \ldots = \rho_w = 1 \\ i_1 < \ldots < i_w \end{array}\right\} \text{ and } \left\{\begin{array}{l} \rho_{w+1} = \ldots = \rho_f = 2 \\ i_{w+1} < \ldots < i_f \end{array}\right.$$

Hence their number N_w equals

$$N_w = \binom{n}{w} \cdot \binom{n}{f - w} = \binom{n}{w} \cdot \binom{n}{\overline{w}} \tag{3.6}$$

where $d = n - f$ and $\overline{w} = w + d$. According to (3.5) the space Ω breaks up into subspaces $\Pi^*{}_u$ of dimensionality $v + 1$ in each of which $B^*(t)$ induces the transformation $B_u(t)$. Every one of these $g^*{}_u$ subspaces $\Pi^*{}_u$, therefore, contributes exactly *one* coordinate of signature w to Ω provided $u \leq w \leq u + v = f - u$. This simple argument yields the recursive formula

$$N_w = \Sigma g^*{}_u$$

where u ranges over all integers satisfying the inequalities $u \geq 0$ and $u \leq w$, $u \leq f - w$. Consequently

$$g^*{}_u = N_u - N_{u-1} \qquad (0 \leq u \leq \tfrac{1}{2}f). \tag{3.7}$$

Put $\overline{u} = d + u$ so that $v = n - u - \overline{u}$. Now (1.6) readily follows from (3.6) and (3.7), and one sees from this explicit expression that $g^*{}_u$ is positive

provided $u \geq 0$, $\bar{u} \geq 0$ and $u + \bar{u} \leq n$. The range of the valences v actually occurring in the decomposition of $\mathfrak{B}^*$ is thus circumscribed by the relations

$$v \geq 0, \quad v \leq n \pm d, \quad v \equiv n \pm d \ (\mathrm{mod}\ 2).^{3}$$

$\mathfrak{B}^*$ serves merely as a jumping board for $\mathfrak{A}^*$. But since every A^* commutes with all the transformations B^* of $\mathfrak{B}^*$ the decomposition (1.5) of the generic matrix A^* of $\mathfrak{A}^*$ is now inferred from Lemma 1. A definite decomposition according to valences is thus obtained, and for physics this is the most essential result. However, as long as we have not yet convinced ourselves that $\mathfrak{A}^*$ is not only contained in, but identical with, the commutator algebra of $\mathfrak{B}^*$, completeness for the decomposition (1.5) is not ensured. In order to settle this point (5) one first has to prove that the only operators in P that commute with the symmetric transformations B are the symmetry operators (4).

4. Symmetric transformations and permutations

Our present object is the space $\Sigma = \Sigma_{n,f}$ of the n-dimensional tensors $\eta(i_1 \ldots i_f)$ of rank f. The permutations $\mathbf{p}$ and any linear combinations of them, $\mathbf{a} = \sum_p a(p)\mathbf{p}$, are linear operators in Σ, $\eta' = \mathbf{a}\eta$, which commute with all the symmetric linear transformations $\dot{\eta} = A\eta$,

$$\dot{\eta}(i_1 \ldots i_f) = \sum_k a(i_1 \ldots i_f\,; k_1 \ldots k_f) \cdot \eta(k_1 \ldots k_f).$$

We introduce the symmetry quantities $\underset{.}{a} = \sum_p a(p)p$ (with arbitrary numbers $a(p)$ as coefficients)[4] quite independently from their usage as operators in Σ. They form an abstract algebra of order $f!$, the "group ring of the symmetric group."

Let η be a tensor and $i_1, \ldots, i_f$ a given sequence of integers from the interval $1 \leq i \leq n$. We consider the $f!$ numbers $\mathbf{p}\eta(i_1 \ldots i_f) = x(p)$ as the coefficients of a symmetry quantity $\underset{.}{x} = \sim\eta(i_1 \ldots i_f)$. The tensor equation $\eta' = \mathbf{a}\eta$ is equivalent with $\sim\eta' = (\sim\eta) \cdot \hat{a}$ where $\hat{a}$ is the symmetry quantity with the coefficients $\hat{a}(p) = a(p^{-1})$. Here $\sim\eta$ may be interpreted as the symmetry quantity with the tensorial coefficients $\sim\eta(p) = \mathbf{p}\eta$, or one may replace $\sim\eta$ and $\sim\eta'$ in our equation by the ordinary symmetry quantities $\sim\eta(i_1 \ldots i_f)$ and $\sim\eta'(i_1 \ldots i_f)$ corresponding to any argument combination $i_1, \ldots, i_f$.

The group ring is an $f!$-dimensional vector space. In it we envisage those symmetry quantities $\sim\eta(i_1 \ldots i_f)$ that arise from arbitrary tensors η and arbitrary argument combinations $(i_1, \ldots, i_f)$, and we determine their linear closure $\kappa = \kappa_n$, i.e. the smallest linear subspace that comprises them all.

---

[3]In passing we notice that the order of the algebra $\mathfrak{B}^*$ may now be evaluated as $\Sigma(v + 1)^2$, the sum extending over the non-negative v of the sequence $f', f' - 2, \ldots$ where $f' = \min(n - d,\ n + d) = \min(f, 2n - f)$, and hence equals $(f' + 1)(f' + 2)(f' + 3)/1 \cdot 2 \cdot 3$. It should be easily possible to confirm this directly.

[4]The dot under a letter merely serves to indicate that it stands for a symmetry quantity.

Let γ_s $(s = 1,2, \ldots, n')$ be a basis for the space Σ. Then the elements $\underset{\cdot}{x}$ of κ are given by the equation

$$\underset{\cdot}{x} = \sum_{s;i} \xi_s(i_1 \ldots i_f) \cdot \sim\gamma_s(i_1 \ldots i_f) \tag{4.1}$$

where the $\xi_s(i_1 \ldots i_f)$ are arbitrary coefficients. Write more explicitly

$$x(p) = \sum_{s;i} \xi_s(i_1 \ldots i_f) \cdot \mathbf{p}\gamma_s(i_1 \ldots i_f) = \sum_{s;i} \mathbf{p}^{-1}\xi_s(i_1 \ldots i_f) \cdot \gamma_s(i_1 \ldots i_f),$$

hence

$$\hat{x} = \sum_{s;i} \gamma_s(i_1 \ldots i_f) \cdot \sim\xi_s(i_1 \ldots i_f). \tag{4.2}$$

Since $\gamma'_s = \mathbf{a}\gamma_s$ implies $\sim\gamma'_s = (\sim\gamma_s)\cdot\hat{a}$ one sees that $\underset{\cdot}{x}\hat{a}$ lies in κ if $\underset{\cdot}{x}$ does; κ is therefore not only an algebra, but even a right-ideal. But in (4.2) one may consider ξ_s as a tensor and the $\gamma_s(i_1 \ldots i_f)$ as coefficients; consequently $\hat{x}$ lies in κ if $\underset{\cdot}{x}$ does, and thus κ is also a left-ideal. Introduce $\xi'_s = \mathbf{a}\xi_s$; then (4.2) yields

$$\hat{x}\cdot\hat{a} = \sum_{s;i} \gamma_s(i_1 \ldots i_f) \cdot \sim\xi'_s(i_1 \ldots i_f),$$

$$\underset{\cdot}{a}\cdot\underset{\cdot}{x} = \sum_{s;i} \xi'_s(i_1 \ldots i_f) \cdot \sim\gamma_s(i_1 \ldots i_f). \tag{4.3}$$

As a left-ideal κ has a generating idempotent $\underset{\cdot}{e}$. This means that $\underset{\cdot}{z}\underset{\cdot}{e}$ is in κ whatever the symmetry quantity $\underset{\cdot}{z}$, and if $\underset{\cdot}{z}$ lies in κ then $\underset{\cdot}{z} = \underset{\cdot}{z}\underset{\cdot}{e}$. Similar statements hold for multiplication by $\hat{e}$ on the left. The ensuing equations $\hat{e} = \hat{e}\cdot\underset{\cdot}{e}$ and $\underset{\cdot}{e} = \hat{e}\cdot\underset{\cdot}{e}$ show that $\underset{\cdot}{e} = \hat{e}$. Every tensor η satisfies the equation $\mathbf{e}\eta = \eta$.

One more fact about κ is of importance. Introduce as the trace $\mathrm{tr}(\underset{\cdot}{a})$ of a symmetry quantity $\underset{\cdot}{a}$ the coefficient $a(\mathbf{1})$ corresponding to the identical permutation $\mathbf{1}$. The scalar product $\mathrm{tr}(\underset{\cdot}{a}\underset{\cdot}{b}) = \sum_p a(p^{-1})\cdot b(p)$ is clearly a symmetric and non-degenerate bilinear form of the two arbitrary symmetry quantities $\underset{\cdot}{a}$ and $\underset{\cdot}{b}$. This non-degeneracy is preserved under restriction to κ; i.e. an $\underset{\cdot}{a} \epsilon \kappa$ such that $\mathrm{tr}(\underset{\cdot}{a}\underset{\cdot}{b}) = 0$ for every $\underset{\cdot}{b} \epsilon \kappa$ is necessarily zero. Indeed let $\underset{\cdot}{z}$ be an arbitrary symmetry quantity; then $\underset{\cdot}{b} = \underset{\cdot}{z}\underset{\cdot}{e}$ is in κ, hence $\mathrm{tr}(\underset{\cdot}{a}\underset{\cdot}{z}\cdot\underset{\cdot}{e}) = 0$. But with $\underset{\cdot}{a}$ also $\underset{\cdot}{a}\underset{\cdot}{z}$ lies in κ, therefore $\underset{\cdot}{a}\underset{\cdot}{z}\cdot\underset{\cdot}{e} = \underset{\cdot}{a}\underset{\cdot}{z}$. Thus our equation turns into $\mathrm{tr}(\underset{\cdot}{a}\underset{\cdot}{z}) = 0$ for every $\underset{\cdot}{z}$, and that implies $\underset{\cdot}{a} = 0$.

THEOREM I. *The symmetry quantities* $\mathbf{a}$ *if interpreted as operators in* Σ *are the only ones that commute with all symmetric transformations* A. *The symmetry quantity* $\underset{\cdot}{a}$ *expressing such an operator can be uniquely normalized by requiring* $\underset{\cdot}{a}$ *to lie in* κ.

Proof (cf. GQ, pp. 266-267).[5] Let L be a linear operator in Σ, $\eta \to L\eta$,

[5]By using deeper algebraic resources than we care to employ in this elementary approach, Theorem I could be obtained as an immediate consequence of the following two facts: (α) Every representation $\underset{\cdot}{a} \to \mathbf{a}$ of the group ring of the symmetric group breaks up into irreducible parts (is "fully reducible"); (β) A fully reducible matric algebra coincides with the commutator algebra of its commutator algebra (R. Brauer).—Another variant: Explicit construction by means of Young's symmetry operators shows that the inequivalent irreducible parts of the representation $a \to \mathbf{a}$ are *absolutely* irreducible and inequivalent, and consequently (α) yields a *complete* decomposition. With this additional knowledge (β) can be replaced by the trivial fact that complete decomposition of a matric algebra implies its identity with the commutator algebra of its commutator algebra.

commuting with all symmetric A. Let $L\gamma_s = \beta_s$, and with the same coefficients $\xi_s(i_1 \ldots i_f)$ as in (4.1) form

$$\underset{\cdot}{y} = \sum_{s;i} \xi_s(i_1 \ldots i_f) \cdot \sim\beta_s(i_1 \ldots i_f).$$

I am going to show that the equation $\underset{\cdot}{x} = 0$ for the arbitrary coefficients $\xi_s(i_1 \ldots i_f)$ implies $\underset{\cdot}{y} = 0$. Let η be any tensor and set

$$\theta = \sum_p x(p^{-1}) \cdot \mathbf{p}\eta, \qquad \tilde{\theta} = \sum_p y(p^{-1}) \cdot \mathbf{p}\eta.$$

Then

$$\theta(i_1 \ldots i_f) = \sum_s \sum_k a_s(i_1 \ldots i_f; k_1 \ldots k_f) \cdot \gamma_s(k_1 \ldots k_f),$$

$$\tilde{\theta}(i_1 \ldots i_f) = \sum_s \sum_k a_s(i_1 \ldots i_f; k_1 \ldots k_f) \cdot \beta_s(k_1 \ldots k_f)$$

where

$$a_s(i_1 \ldots i_f; k_1 \ldots k_f) = \sum_p \mathbf{p}\eta(i_1 \ldots i_f) \cdot \mathbf{p}\xi_s(k_1 \ldots k_f)$$

is clearly the matrix of a symmetric operator A_s in Σ. As A_s commutes with L we conclude that $\tilde{\theta} = L\theta$. Consequently $\theta = 0$ implies $\tilde{\theta} = 0$, and $\underset{\cdot}{x} = 0$ implies $\sum_p y(p^{-1}) \cdot \mathbf{p}\eta(i_1 \ldots i_f) = 0$, or tr $(\underset{\cdot}{y}\underset{\cdot}{y}^*) = 0$ for every $\underset{\cdot}{y}^* \in \kappa$. The quantity $\underset{\cdot}{y}$ itself is in κ, and hence the last equation forces $\underset{\cdot}{y}$ to vanish.

This settled, one concludes that the correspondence $\underset{\cdot}{x} \to \underset{\cdot}{y} = R\underset{\cdot}{x}$ defines a linear mapping R of κ into itself. Formula (4.3) and its parallel

$$\underset{\cdot}{a} \cdot \underset{\cdot}{y} = \sum_{s;i} \xi'_s(i_1 \ldots i_f) \cdot \sim\beta_s(i_1 \ldots i_f)$$

prove the mapping R to be a similarity; i.e. it carries $\underset{\cdot}{a}\underset{\cdot}{x}$ into $\underset{\cdot}{a}\underset{\cdot}{y}$ whatever $\underset{\cdot}{a}$. Replace $\underset{\cdot}{x}$ and $\underset{\cdot}{a}$ by $\underset{\cdot}{e}$ and $\underset{\cdot}{x}$. Setting $R\underset{\cdot}{e} = \underset{\cdot}{\mathfrak{a}}$ one finds that $\underset{\cdot}{x} = \underset{\cdot}{x}\underset{\cdot}{e}$ goes into $\underset{\cdot}{x} \cdot R\underset{\cdot}{e} = \underset{\cdot}{x}\underset{\cdot}{\mathfrak{a}}$. This statement is equivalent with the n^f equations $\beta_s = \mathbf{a}\gamma_s$, or $L\eta = \mathbf{a}\eta$ for every tensor η. The symmetry quantities $\underset{\cdot}{\mathfrak{a}}$ and $\underset{\cdot}{a}$ lie in κ.

5. The reciprocity of $\mathfrak{A}^*$ and $\mathfrak{B}^*$

In this section ν is not assumed to have the special value 2.

THEOREM II. *$\mathfrak{A}^*$ is the commutator algebra of $\mathfrak{B}^*$.*

Proof. Let

$$C = \| c(i_1\rho_1, \ldots, i_f\rho_f; k_1\sigma_1, \ldots, k_f\sigma_f) \|$$

be the matrix of any linear transformation in Ω in the unique normalization established by Lemma 3. Hence C is antisymmetric in the f pairs $(i\rho)$, antisymmetric in the f pairs $(k\sigma)$, and thereby symmetric in the f quadruples $(i\rho, k\sigma)$. Let $X = \| x(\rho_1 \ldots \rho_f; \sigma_1 \ldots \sigma_f) \|$ be symmetric in the f pairs $(\rho\sigma)$. Then CX with the components

$$\sum_\tau c(i_1\rho_1, \ldots, i_f\rho_f; k_1\tau_1, \ldots, k_f\tau_f) \cdot x(\tau_1 \ldots \tau_f; \sigma_1 \ldots \sigma_f)$$

is certainly antisymmetric in the pairs $(i\rho)$, and since it is symmetric in the quadruples $(i\rho, k\sigma)$ it is also antisymmetric in the pairs $(k\sigma)$. The same is true for XC. Our hypothesis demands that CX and XC coincide as operators

in Ω. Hence their matrices in normalized form must be identical. For fixed $i_1 \dots i_f ; k_1 \dots k_f$ the coefficients

$$c(\rho_1 \dots \rho_f ; \sigma_1 \dots \sigma_f) = c(i_1\rho_1, \dots, i_f\rho_f ; k_1\sigma_1, \dots, k_f\sigma_f)$$

form a matrix $\| c(\rho_1 \dots \rho_f ; \sigma_1 \dots \sigma_f) \|$ in P which may be denoted by $C(i_1 \dots i_f ; k_1 \dots k_f)$. Theorem I when applied to P rather than Σ shows that this transformation is of the form $\sum_p t_p \mathbf{p}$ where

$$t(p) = t_p = t_p(i_1 \dots i_f ; k_1 \dots k_f)$$

are the coefficients of a symmetry quantity

$$t = t(i_1 \dots i_f ; k_1 \dots k_f) \tag{5.1}$$

that lies in $\kappa = \kappa_\nu$. Introduce the transformation

$$T_p = \| t_p(i_1 \dots i_f ; k_1 \dots k_f) \| \tag{5.2}$$

in Σ. Our result may then be written in the form

$$C = \sum_p (T_p \times \mathbf{p}),$$

the cross indicating the Kronecker product of a matrix in Σ (first factor) and a matrix in P (second factor). If we are not afraid of making use of a symmetry quantity T whose coefficients are the matrices T_p in Σ we can express the fact that each t lies in κ_ν by the equations

$$Te = eT = T, \tag{5.3}$$

$e = e_\nu$ being the generating idempotent of $\kappa = \kappa_\nu$.

C is antisymmetric in the pairs $(k\sigma)$. Hence

$$C(\mathbf{q} \times \mathbf{q}) = \delta_q \cdot C \tag{5.4}$$

for any permutation q. It is antisymmetric in the pairs $(i\rho)$; hence also

$$(\mathbf{q} \times \mathbf{q})C = \delta_q \cdot C. \tag{5.5}$$

In more explicit form (5.4) reads

$$\sum_p \{ T_p \mathbf{q} \times \mathbf{pq} \} = \delta_q \cdot \sum_p \{ T_p \times \mathbf{p} \}$$

or

$$\sum_p \{ T_{pq^{-1}} \mathbf{q} \times \mathbf{p} \} = \delta_q \cdot \sum_p \{ T_p \times \mathbf{p} \}. \tag{5.6}$$

In order to avoid confusion use for the moment

$$Q = \| q(i_1 \dots i_f ; k_1 \dots k_f) \|$$

as a notation for the linear transformation $\mathbf{q}$ in Σ and its matrix. Set $T'_p = T_p Q$,

$$t'_p(i_1 \dots i_f ; k_1 \dots k_f)$$
$$= \sum_l t_p(i_1 \dots i_f ; l_1 \dots l_f) \cdot q(l_1 \dots l_f ; k_1 \dots k_f).$$

Given a combination $(i_1 \dots i_f ; k_1 \dots k_f)$, the symmetry quantity t' with the coefficients $t'(p) = t'_p(i_1 \dots i_f ; k_1 \dots k_f)$ lies in κ_ν because all the quantities $t(i_1 \dots i_f ; l_1 \dots l_f)$ do. (This is true for any linear transformation Q in Σ. What holds for $T'_p = T_p Q$ holds likewise for $T''_p = QT_p$.) For a fixed permutation q the numbers $t^*(p) = t'(pq^{-1})$ are the coefficients of the symmetry quantity $t^* = t'q$. $\| t^*_p(i_1 \dots i_f ; k_1 \dots k_f) \|$ is the matrix $T'_{pq^{-1}} = T_{pq^{-1}} Q$. Hence (5.6) states that t^* and $\delta_q \cdot t$ coincide as symmetry operators in P.

But $\underset{\cdot}{t}^* = \underset{\cdot}{t}'\underset{\cdot}{q}$ lies in κ, because $\underset{\cdot}{t}'$ does; coincidence as operators in P, therefore, implies identity of the symmetry quantities themselves, $\underset{\cdot}{t}^* = \delta_q \cdot \underset{\cdot}{t}$ or

$$T_{pq^{-1}}\mathbf{q} = \delta_q \cdot T_p.$$

Setting $q = p$, $T_1 = A$, one finds

$$T_p = \delta_p \cdot A\mathbf{p}.$$

In the same manner (5.5) leads to

$$T_p = \delta_p \cdot \mathbf{p}A.$$

The transformation A in Σ thus commutes with the permutation operators $\mathbf{p}$ in the same space and is therefore symmetric. Because of the antisymmetry of $\psi(i_1\rho_1, \ldots, i_f\rho_f)$ in the pairs $(i\rho)$ the equation

$$\psi' = C\psi = \textstyle\sum_p (T_p \times \mathbf{p})\psi$$

may be written as

$$\psi' = \textstyle\sum_p \delta_p \cdot (T_p \mathbf{p}^{-1} \times I)\psi = f!(A \times I)\psi$$

where I stands for identity, and thus Theorem II is proved.

The normalizing condition (5.3) takes on the form

(5.7) $$\check{e}A = A\check{e} = A,$$

$\check{e}$ being the idempotent with the coefficients $\delta_p \cdot e(p) = \delta_p \cdot e(p^{-1})$. This, however, is no surprise. As a matter of fact, A induces the same transformation A^* in Ω as $\check{e}A\check{e}$, and hence, whether or not A satisfies (5.7), it can always be so modified as to fulfil that relation, without change in the corresponding A^*.

Application of Theorem II to $\nu = 2$ shows that the decomposition (1.5) by valences is complete.

Institute for Advanced Study
Princeton, New Jersey

ORTHOGONAL MATRICES IN FOUR-SPACE

C. C. MacDUFFEE

Every proper orthogonal matrix A can be written

$$A = e^{Q}$$

where Q is a skew matrix [6], and conversely every such matrix A is orthogonal. It is also known that every proper orthogonal transformation in real Euclidean four-space may be characterized in term of quaternions [1, 3] by the equation

$$x' = axb, \qquad \mathrm{N}a = \mathrm{N}b = 1.$$

Here the quaternion

$$x = x_0 + x_1 i + x_2 j + x_3 k$$

determines with the origin a vector having the coordinates (x_0, x_1, x_2, x_3). The relationship between these two representations was clearly shown by Murnaghan [5].

The present paper employs the first and second regular representations of quaternions by matrices in place of Murnaghan's "special matrices," with the result that known properties of the regular representations can be applied directly to this problem. Incidentally an easy method not using infinite series is found for finding the skew matrix Q when the orthogonal matrix A is given.

1. The first and second regular representations of the real quaternion

$$\alpha = a_0 + a_1 i + a_2 j + a_3 k$$

are, respectively,

$$R(\alpha) = a_0 I + a_1 R_1 + a_2 R_2 + a_3 R_3, \qquad S(\alpha) = a_0 I + a_1 S_1 + a_2 S_2 + a_3 S_3$$

where

$$R_1 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad R_2 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \quad R_3 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

(1)

$$S_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \quad S_3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}.$$

Let S^T denote the transpose of S. The six matrices R_1, R_2, R_3, S_1^T, S_2^T, S_3^T are all skew and are linearly independent. The most general 4 by 4 skew matrix

Received February 25, 1948.

$$(2) \qquad Q = \begin{bmatrix} 0 & q_{01} & q_{02} & q_{03} \\ -q_{01} & 0 & q_{12} & -q_{31} \\ -q_{02} & -q_{12} & 0 & q_{23} \\ -q_{03} & q_{31} & -q_{23} & 0 \end{bmatrix}$$

is therefore a linear combination of them. In fact

$$Q = -\tfrac{1}{2}(q_{01} + q_{23})R_1 - \tfrac{1}{2}(q_{02} + q_{31})R_2 - \tfrac{1}{2}(q_{03} + q_{12})R_3$$
$$-\tfrac{1}{2}(q_{01} - q_{23})S_1^T - \tfrac{1}{2}(q_{02} - q_{31})S_2^T - \tfrac{1}{2}(q_{03} - q_{12})S_3^T.$$

Note the analogy of the q's to Plücker line coordinates [2].

If we let $-\tfrac{1}{2}(q_{01} + q_{23}) = r_1$, $-\tfrac{1}{2}(q_{01} - q_{23}) = s_1$ etc., we may write

$$\rho = r_1 i + r_2 j + r_3 k, \qquad \sigma = s_1 i + s_2 j + s_3 k.$$

That is, every skew matrix can be written

$$Q = R(\rho) + S^T(\sigma),$$

where ρ and σ are pure quaternions. Therefore ρ satisfies the quadratic equation

$$(3) \qquad x^2 + \mathrm{N}\rho = 0, \qquad \mathrm{N}\rho = r_1^2 + r_2^2 + r_3^2,$$

and similarly for σ.

The matrix e^Q is defined as a power series which converges for every matrix Q. In every associative algebra, every matrix of the first regular representation is commutative with the transpose of every matrix of the second regular representation [4]. It follows upon multiplying power series that

$$e^Q = e^{R(\rho)}e^{S^T(\sigma)} = e^{S^T(\sigma)}e^{R(\rho)}.$$

Write R for $R(\rho)$. Then

$$e^R = I + \frac{1}{2}R^2 + \frac{1}{4!}R^4 + \cdots + R\left[I + \frac{1}{3!}R^2 + \frac{1}{5!}R^4 + \cdots\right].$$

From (3), $R^2 = -v^2 I$ where $v^2 = \mathrm{N}\rho$, $v \geqq 0$. Hence

$$e^R = \cos v \cdot I + \frac{R}{v}\sin v.$$

If we define the quaternion

$$(4) \qquad \alpha = \cos v + \frac{\rho}{v}\sin v,$$

then clearly

$$R(\alpha) = e^{R(\rho)}, \qquad \mathrm{N}\alpha = 1.$$

By means of (4) every pure quaternion ρ determines a unit quaternion α and vice versa. Similarly

$$e^{S^T(\sigma)} = S^T(\beta), \qquad \mathrm{N}\beta = 1.$$

We have proved

THEOREM 1. *Every real proper 4 by 4 orthogonal matrix can be written*

$$A = R(\alpha)\cdot S^T(\beta) = S^T(\beta)\cdot R(\alpha)$$

where α and β are unit quaternions. Every such product is orthogonal and proper.

Let us assume a second such representation,

$$A = R(\gamma)\cdot S^T(\delta), \qquad N\gamma = N\delta = 1.$$

Then

$$R^{-1}(\gamma)\cdot R(\alpha) = S^T(\delta)\cdot S^{-T}(\beta), \quad R(\gamma^{-1}\alpha) = S^T(\beta^{-1}\delta).$$

The skew components of these matrices vanish, since the skew matrices in (1) are linearly independent. Thus

$$R(\gamma^{-1}\alpha) = S^T(\beta^{-1}\delta) = kI, \qquad k \text{ real},$$

so that $\alpha = k\gamma, \delta = k\beta$. Since $N\alpha = N\gamma = 1, Nk = k^2 = 1, k = \pm 1$. We have

THEOREM 2. *The pair of quaternions* α, β *of Theorem* 1 *is unique except that it may be replaced by* $-\alpha$, $-\beta$.

2. The unit quaternion

$$\alpha = a_0 + a_1 i + a_2 j + a_3 k, \qquad N\alpha = 1, \tag{5}$$

satisfies the quadratic equation

$$x^2 - 2a_0 x + 1 = 0, \tag{6}$$

whose roots are the characteristic roots of $R(\alpha)$. Since the discriminant is $-4(a_1^2 + a_2^2 + a_3^2)$, these characteristic roots are real only if $\alpha = \pm 1$. That is, unless the orthogonal matrix $R(\alpha)$ is $\pm$ I, the orthogonal transformation which it defines leaves no vector through the origin invariant. But if v is any vector through the origin, the plane of vectors $k_1 v + k_2 R(\alpha)\cdot v$ is invariant. For by (6)

$$R(\alpha)[k_1 v + k_2 R(\alpha)\cdot v] = -k_2 v + (k_1 + 2a_0 k_2)R(\alpha)\cdot v.$$

Thus $R(\alpha)$ is the matrix of a left Clifford translation.

Coxeter [1] has shown that in quaternion coordinates the left Clifford translation is given by

$$x' = \alpha x, \qquad N\alpha = 1,$$

where α is given by (5), and

$$x = x_0 + x_1 i + x_2 j + x_3 k.$$

Upon multiplying out and equating the coefficients of 1, i, j and k, we have

$$\begin{aligned}
x'_0 &= a_0x_0 - a_1x_1 - a_2x_2 - a_3x_3, \\
x'_1 &= a_1x_0 + a_0x_1 - a_3x_2 + a_2x_3, \\
x'_2 &= a_2x_0 + a_3x_1 + a_0x_2 - a_1x_3, \\
x'_3 &= a_3x_0 - a_2x_1 + a_1x_2 + a_0x_3.
\end{aligned}$$

If we denote by v the column vector with components x_0, x_1, x_2, x_3, this may be written

$$v' = R(\alpha)\cdot v, \qquad N\alpha = 1.$$

In the same notation the right Clifford translations may be written

$$v' = S^T(\beta)\cdot v, \qquad N\beta = 1.$$

3. It has been shown that if A is proper orthogonal,

$$A = R(\alpha)\cdot S^T(\beta),$$

where α is given by (5) and β is given similarly. We shall show how α and β can be determined from A. From (1)

$$A = a_0b_0I + \sum_{i,j=1}^{3} a_ib_jR_iS_j^T + a_0 \sum_{i=1}^{3} b_iS_i^T + b_0 \sum_{j=1}^{3} a_jR_j. \tag{7}$$

Since R_i and S_j^T are both skew and commutative, their product is symmetric. Thus the first ten terms above are symmetric and the last six are skew. Hence the unique skew part of A is

$$\tfrac{1}{2}(A - A^T) = a_0[b_1S_1^T + b_2S_2^T + b_3S_3^T] + b_0[a_1R_1 + a_2R_2 + a_3R_3].$$

Since the R_i and S_j^T are linearly independent, we can determine uniquely the numerical values of

$$a_0b_1,\ a_0b_2,\ a_0b_3,\ b_0a_1,\ b_0a_2,\ b_0a_3.$$

With the aid of the relations

$$a_0^2 + a_1^2 + a_2^2 + a_3^2 = 1, \qquad b_0^2 + b_1^2 + b_2^2 + b_3^2 = 1,$$

we obtain quadratic equations for a_0^2 and b_0^2 and hence the values of the eight a_i and b_j. It is known from Theorem 2 that just two sets of values can satisfy (7).

When α and β are known, ρ and σ can be found from (4), and then Q from (2).

REFERENCES

[1] H. S. M. Coxeter, *American Mathematical Monthly*, vol. 53 (1946), 136-146.

[2] H. S. M. Coxeter, *Non-Euclidean Geometry* (Toronto, 1942), 151-153.

[3] E. Guth, *Anzeiger, Akademie der Wissenschaften in Wien, Mathematisch-naturwissenschaftliche Klasse*, July 6, 1933, Nr. 18.

[4] C. C. MacDuffee, *Monatshefte für Mathematik und Physik*, vol. 48 (1939), 294.

[5] F. D. Murnaghan, *Scripta Mathematica*, vol. 10 (1944), 37-49.

[6] J. H. M. Wedderburn, *Ann. of Math.*, vol. 23 (1921), 134.

The University of Wisconsin

ON CONJUGATE CONVEX FUNCTIONS

W. FENCHEL

1. Since the classical work of Minkowski and Jensen it is well known that many of the inequalities used in analysis may be considered as consequences of the convexity of certain functions. In several of these inequalities pairs of "conjugate" functions occur, for instance pairs of powers with exponents a and α related by $1/a + 1/\alpha = 1$. A more general example is the pair of positively homogeneous convex functions defined by Minkowski and known as the distance (or gauge) function and the function of support of a convex body. The purpose of the present paper is to explain the general (by the way rather elementary) idea underlying this correspondence. Subjected to a more precise formulation the result is the following:

To each convex function $f(x_1, \ldots, x_n)$ defined in a convex region G and satisfying certain conditions of continuity there corresponds in a unique way a convex region Γ and a convex function $\phi(\xi_1, \ldots, \xi_n)$ defined in Γ and with the same properties such that

$$x_1\xi_1 + \ldots + x_n\xi_n \leqq f(x_1, \ldots, x_n) + \phi(\xi_1, \ldots, \xi_n), \tag{1}$$

for all points $(x_1, \ldots, x_n)$ in G and all points $(\xi_1, \ldots, \xi_n)$ in Γ. The inequality is exact in a sense explained below. The correspondence between G, f and Γ, ϕ is symmetric, and the functions f and ϕ are called conjugate.[1]

The hypersurfaces $y = f(x_1, \ldots, x_n)$ and $\eta = \phi(\xi_1, \ldots, \xi_n)$ correspond to each other in the polarity with respect to the paraboloid

$$2y = x_1^2 + \ldots + x_n^2.$$

Let $F(x)$ be strictly increasing for $x \geqq 0$. Then $f(x) = \int_0^x F(x)dx$ is convex, and its conjugate function is $\phi(\xi) = \int_0^\xi \Phi(\xi)d\xi$ where $\Phi(\xi)$ is the inverse function of $F(x)$. The inequality (1) for $n = 1$ therefore yields the well-known inequality of W. H. Young[2]

$$x\xi \leqq \int_0^x F(x)dx + \int_0^\xi \Phi(\xi)d\xi.$$

(1) may thus be considered as a generalization of this inequality.

Received March 24, 1948.

[1]The case $n = 1$ has been considered by S. Mandelbrojt [3] under the assumption that the ranges G and Γ are identical with the entire axis $-\infty < x < \infty$. This, however, is incompatible with the complete reciprocity between f and ϕ which will appear from an example given below. Mandelbrojt's formulation of the theorem is thus not quite correct due to the fact that the least upper bounds occurring in it may be infinite.

[2]See e.g. [2] p. 111.

If $f(x_1, \ldots, x_n)$ is positively homogeneous of degree one, then G is the entire space $x_1, \ldots, x_n$ while Γ is closed and bounded, and $\phi(\xi_1, \ldots, \xi_n)$ is identically zero. In this case (1) expresses that $f(x_1, \ldots, x_n)$ is the function of support of the convex body Γ.[3]

2. The euclidean spaces with coordinates $x_1, \ldots, x_n$ and $x_1, \ldots, x_n, y$ will be denoted by R^n and R^{n+1} respectively, points and vectors in these spaces by x and x, y respectively. Furthermore we write

$$x' + x'' = (x'_1 + x''_1, \ldots, x'_n + x''_n), \quad \lambda x = (\lambda x_1, \ldots, \lambda x_n),$$

$$\Sigma x\xi = x_1\xi_1 + \ldots + x_n\xi_n.$$

θ will always denote a number in the interval $0 < \theta < 1$.

The point set G of R^n is supposed to be convex, i.e. if x' and x'' belong to G, the whole segment $(1 - \theta)x' + \theta x''$ belongs to G. But G need neither be closed nor open nor bounded. The interior points of segments belonging to G are shortly called the interior points of G. All other points of accumulation of G, belonging to G or not, will be called the boundary or extreme points of G.

A function $f(x)$ defined in G is called convex if

$$f\left((1 - \theta)x' + \theta x''\right) \leqq (1 - \theta)f(x') + \theta f(x'') \tag{2}$$

for any two points x' and x'' of G and all θ. It is well known that this implies that $f(x)$ is continuous at the interior points of G. For our purpose we have also to consider the behaviour of $f(x)$ at the boundary points. Let x^* be a boundary point of G. For functions of one variable $\lim_{x \to x^*} f(x)$ exists or is ∞. But this is not necessarily the case for functions of several variables. If x^* belongs to G the only general conclusion to be drawn from (2) is that

$$\underline{\lim}_{x \to x^*} f(x) \leqq f(x^*); \tag{3}$$

for, from

$$f\left((1 - \theta)x + \theta x^*\right) \leqq (1 - \theta)f(x) + \theta f(x^*)$$

it follows that

$$\underline{\lim}_{x \to x^*} f(x) \leqq \underline{\lim}_{\theta \to 1} f\left((1 - \theta)x + \theta x^*\right) \leqq f(x^*),$$

and (2) remains valid if $f(x^*)$ is replaced by any other value satisfying (3).

If necessary, we now change G and f by adding to G all those boundary points x^* not yet belonging to G for which $\underline{\lim}_{x \to x^*} f(x)$ is finite and by defining f at these and at the boundary points previously belonging to G by

$$f(x^*) = \underline{\lim}_{x \to x^*} f(x). \tag{4}$$

The new G and the function f obtained in this way are obviously again convex; for, let x' and x'' be arbitrary points of the new G and $x'_{(\nu)}$ and $x''_{(\nu)}$, $\nu = 1, 2, \ldots,$

---

[3] See e.g. [1] p. 23-24.

sequences of interior points of G such that

$$x'_{(\nu)} \to x',\ x''_{(\nu)} \to x'',\ f(x'_{(\nu)}) \to f(x'),\ f(x''_{(\nu)}) \to f(x''),$$

then we get from

$$f\big((1-\theta)x'_{(\nu)} + \theta x''_{(\nu)}\big) \leqq (1-\theta)f(x'_{(\nu)}) + \theta f(x''_{(\nu)})$$

for $\nu \to \infty$

$$\varliminf_{x \to (1-\theta)x' + \theta x''} f(x) \leqq \varliminf_{\nu \to \infty} f\big((1-\theta)x'_{(\nu)} + \theta x''_{(\nu)}\big) \leqq (1-\theta)f(x') + \theta f(x''),$$

which shows that $(1-\theta)x' + \theta x''$ belongs to G and that (2) is valid, as the left-hand side is $f\big((1-\theta)x' + \theta x''\big)$.

With (3) in mind we may say that (4) expresses that the functions which will be considered in the following are convex and semi-continuous from below, and G is "closed relative to f," i.e. all boundary points at which $\varliminf f(x)$ is finite belong to G, or in other words, at each boundary point which does not belong to G we have $\lim f(x) = \infty$.

3. The theorem to be proved may now be formulated thus:

Let G be a convex point set in R^n and $f(x)$ a function defined in G convex and semi-continuous from below and such that $\lim_{x \to x^*} f(x) = \infty$ for each boundary point x^* of G which does not belong to G. Then there exists one and only one point set Γ in R^n and one and only one function $\phi(\xi)$ defined in Γ with exactly the same properties as G and $f(x)$ such that

(5) $$\Sigma x\xi \leqq f(x) + \phi(\xi),$$

where to every interior point x of G there corresponds at least one point ξ of Γ for which equality holds.

In the same way G, $f(x)$ correspond to Γ, $\phi(\xi)$.

We define Γ as the set of all points ξ with the property that the function $\Sigma x\xi - f(x)$ is bounded from above in G, and we define $\phi(\xi)$ in Γ as the least upper bound of this function:

$$\phi(\xi) = \underset{x \in G}{\text{l.u.b.}} \big(\Sigma x\xi - f(x)\big).$$

Then (5) is valid. The inequality $\Sigma x\xi - f(x) \leqq s$ or

$$f(x) \geqq \Sigma x\xi - s$$

means that the hyperplane $y = \Sigma x\xi - s$ in R^{n+1} with the normal vector $\xi, -1$ lies nowhere above the hypersurface $y = f(x)$, and $-s$ is the intercept of this hyperplane on the y-axis. It is a well-known fact that there exists at least one hyperplane of support of the convex hypersurface, i.e. a hyperplane which contains at least one point of the hypersurface and lies nowhere above it. This shows that Γ is not empty. Further we see that if there exists a hyperplane of support with the normal vector $\xi, -1$ and if $x^0, f(x^0)$ is a point of contact, then we have

$$\phi(\xi) = \Sigma x^0\xi - f(x^0),$$

and $-\phi(\xi)$ is the y-intercept of this hyperplane. If x^0 is an arbitrary interior point of G, a hyperplane of support through $x^0, f(x^0)$ exists, and this proves the assertion on the equality sign in (5).

It is evident that Γ and $\phi(\xi)$ are convex. In fact, let ξ' and ξ'' be arbitrary points of Γ, then we have for $x \epsilon G$,

$$\Sigma x\xi' - f(x) \leqq \phi(\xi'), \; \Sigma x\xi'' - f(x) \leqq \phi(\xi''),$$

hence

$$\Sigma x((1-\theta)\xi' + \theta\xi'') - f(x) \leqq (1-\theta)\phi(\xi') + \theta\phi(\xi'')$$

which shows that $(1-\theta)\xi' + \theta\xi''$ belongs to Γ and that

$$\phi((1-\theta)\xi' + \theta\xi'') \leqq (1-\theta)\phi(\xi') + \theta\phi(\xi'').$$

Let now ξ^* be a boundary point of Γ and $\xi \epsilon \Gamma$, $x \epsilon G$. Then it follows from (5) that

$$\varliminf_{\xi \to \xi^*} \phi(\xi) \geqq \Sigma x\xi^* - f(x)$$

and this shows on the one hand that $\xi^* \epsilon \Gamma$ if $\varliminf \phi(\xi)$ is finite, i.e. that Γ is closed relative to $\phi(\xi)$, and on the other hand that

$$\varliminf_{\xi \to \xi^*} \phi(\xi) \geqq \phi(\xi^*),$$

i.e. that $\phi(\xi)$ is semi-continuous from below. Hence Γ and ϕ have the same properties as G and f.

4. It remains to be proved that if we start with Γ and $\phi(\xi)$ the same procedure gives G and $f(x)$ again. We have to consider the set G^* of all points x for which $\Sigma\xi x - \phi(\xi)$ is bounded from above in Γ, together with the function

$$f^*(x) = \underset{\xi \epsilon \Gamma}{\text{l.u.b.}} \left(\Sigma\xi x - \phi(\xi)\right)$$

defined in G^*.

If $x \epsilon G$ we get from (5)

$$\Sigma\xi x - \phi(\xi) \leqq f(x) \tag{6}$$

for all $\xi \epsilon \Gamma$, hence $G \subset G^*$ and $f^*(x) \leqq f(x)$ in G. But to an interior point x of G there corresponds a ξ such that equality is valid in (6), which implies $f^*(x) \geqq f(x)$. Hence $f^*(x) = f(x)$ at the interior points of G and, as both functions are convex and semi-continuous from below, also at the boundary points of G.

Let now x^0 be a point of R^n not in G. We have to prove that it does not belong to G^*, i.e. that

$$\underset{\xi \epsilon \Gamma}{\text{l.u.b.}} \left(\Sigma\xi x^0 - \phi(\xi)\right) = \infty. \tag{7}$$

Since the quantity $\Sigma\xi x^0 - \phi(\xi)$ is the y-coordinate of the point at which the hyperplane

$$y = \Sigma\xi x - \phi(\xi)$$

of R^{n+1} intersects the line $x = x^0$ parallel to the y-axis, we have to show that there are hyperplanes below the hypersurface $y = f(x)$ which have arbitrary large intercepts on the line $x = x^0$. Suppose first that x^0 is an exterior point of G. Then there exists a hyperplane H parallel to the y-axis which separates the line $x = x^0$ from G and $y = f(x)$. Consider any hyperplane of support S of $y = f(x)$. Let S turn around the intersection of H and S so that the part lying below $y = f(x)$ moves downwards. Then the point at which S intersects the line $x = x^0$ moves upwards and tends to infinity. Suppose next that x^0 is a boundary point of G but not belonging to G. Then we have $f(x) \to \infty$ for $x \to x^0$. Consider any segment belonging to G and having x^0 as one of its end points. Let x' be a fixed point and x'' a variable point of the segment between

x' and x^0 such that $f(x'') > f(x')$. A plane of support through x'', $f(x'')$ then intersects the line $x = x^0$ at a point the y-coordinate of which is greater than $f(x'')$ and therefore tends to infinity if $x'' \to x^0$. This completes the proof of the theorem.

5. In section 1 it has been asserted that the hypersurfaces $y = f(x)$ and $\eta = \phi(\xi)$ correspond to each other in the polarity with respect to $2y = \Sigma x^2$. This is obviously true in the sense that each of the hypersurfaces is the envelope of the polar hyperplanes of the points of the other. For $y = f(x)$ may be considered as the envelope of the hyperplanes

$$y = \Sigma x\xi - \phi(\xi),$$

where $\xi \epsilon \Gamma$ is the parameter, and the poles of these hyperplanes are the points ξ, $\phi(\xi)$.

6. Suppose now that $y = f(x)$ is strictly convex, i.e. each hyperplane of support contains only one point of $y = f(x)$. Let further $\eta = \phi(\xi)$ satisfy the same condition; for $y = f(x)$ this means that there passes at most one hyperplane of support through a point of $y = f(x)$. Then $f(x)$ has continuous derivatives[4]

$$f_i(x) = \frac{\partial f}{\partial x_i}$$

and we have

$$\xi_i = f_i(x).$$

These relations establish a continuous one to one correspondence between the interior points of G and those of Γ. Solving them with respect to the x we get

$$x_i = \phi_i(\xi)$$

where, for reasons of symmetry, the ϕ_i must be the derivatives of ϕ. From this it is seen that in the case of $n = 1$ the derivatives of two conjugate convex functions are mutually inverse functions. This proves the assertion of section 1 on the inequality of Young. Furthermore we get an explicit expression for $\phi(\xi)$ if $f(x)$ is given, viz.

$$\phi(\xi) = \sum_{i=1}^{n} \xi_i\phi_i(\xi) - f\left(\phi_i(\xi)\right)$$

valid in the interior of Γ. Hence, our correspondence between f and ϕ is the Legendre transformation of the theory of differential equations.

References

[1] T. Bonnesen, W. Fenchel, *Theorie der konvexen Körper* (Berlin, 1934).

[2] G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities* (Cambridge, 1934).

[3] S. Mandelbrojt, "Sur les fonctions convexes," *C.R. Acad. Sci. Paris*, vol. 209 (1939), 977-978.

The Technical University of Denmark
Copenhagen

[4]See [1] p. 23, 26. The argument used there in the case of positively homogeneous convex functions may easily be generalized to the case considered here.

ON THE CRITICAL LATTICES OF ARBITRARY POINT SETS

K. MAHLER

(*Dedicated to J. G. van der Corput*)

In this note, I shall establish necessary and sufficient conditions for the existence of critical lattices of an *arbitrary point set*, and I shall construct a non-trivial example of a point set without any critical lattice. In a previous paper,[1] I proved that *every star body of the finite type possesses at least one critical lattice.*

I.

1. Let S be any point set in n-dimensional Euclidean space R_n. A lattice Λ is called *S-admissible* if no point of Λ, except possibly the origin

$$O = (0, 0, \dots, 0),$$

is an *inner* point of S. Such admissible lattices need not exist, e.g. if S is the whole space R_n; we say in this case that S is of the *infinite type*, and put

$$\Delta(S) = \infty.$$

If there are admissible lattices, S is called of the *finite type*. We then form the lower bound

$$\Delta(S) = \text{l.b. } d(\Lambda)$$

of the determinants $d(\Lambda)$ of all S-admissible lattices, and call this the *minimum determinant of S*. In the special case that

$$\Delta(S) = 0,$$

there exist S-admissible lattices of arbitrarily small determinant, and S is called of the *zero type*; e.g. the null set has this property.

2. A lattice Λ is called a *critical lattice of S* if

(*a*) Λ is S-admissible, and

(*b*) $d(\Lambda) = \Delta(S)$.

It is clear from the definitions just given that S cannot have a critical lattice if it is of the infinite or the zero types. For there are no S-admissible lattices in the first case; in the second case, the lower bound is not attained since every lattice is of positive determinant.

In the remaining case, when

$$0 < \Delta(S) < \infty, \tag{1}$$

the following criterion holds.

Received April 30, 1948.

[1] "On lattice points in n-dimensional star bodies, I," Proc. Royal Soc., A, 187 (1946), 151-187. The letters LP will be used to mark references to this paper.

THEOREM 1. *Let S be a point set in R_n satisfying* (1). *Then S possesses at least one critical lattice, if and only if there exists a bounded*[2] *infinite sequence of S-admissible lattices*

$$\Lambda_1, \Lambda_2, \Lambda_3, \ldots$$

such that

$$\lim_{r\to\infty} d(\Lambda_r) = \Delta(S). \tag{2}$$

Proof. (i) If there exists a critical lattice Λ of S, then the infinite sequence of lattices

$$\Lambda, \Lambda, \Lambda, \ldots$$

has the required properties.

(ii) Assume that

$$\Lambda_1, \Lambda_2, \Lambda_3, \ldots$$

is a bounded infinite sequence of S-admissible lattices satisfying (2). We may then select[3] an infinite subsequence

$$\Lambda_{r_1}, \Lambda_{r_2}, \Lambda_{r_3}, \ldots \quad (r_1 < r_2 < r_3 < \ldots)$$

tending to a limit, the lattice Λ say. By the continuity of the determinant,

$$d(\Lambda) = \lim_{k\to\infty} d(\Lambda_{r_k}) = \Delta(S).$$

The assertion is therefore proved if we can show that Λ is S-admissible, hence critical. If Λ were not S-admissible, there would be a point $P \neq O$ of Λ which is an *inner* point of S. There exists then a neighbourhood of P consisting only of inner points of S. Since the lattices Λ_{r_k} tend to Λ, this neighbourhood contains a point of Λ_{r_k} for all sufficiently large indices k, contrary to the assumption that Λ_{r_k} is S-admissible.

3. Two special cases of Theorem 1 are of particular interest.

THEOREM 2. *If the point set S is of the finite type, and if O is an inner point of S, then S possesses at least one critical lattice.*

Proof. Choose an arbitrary infinite sequence of S-admissible lattices

$$\Lambda_1, \Lambda_2, \Lambda_3, \ldots$$

satisfying (2). Then this sequence is bounded since none of its points lie in a sufficiently small neighbourhood of O. The assertion follows therefore immediately from Theorem 1.

THEOREM 3. *If the point set S is bounded and not of the zero type, then it possesses at least one critical lattice.*

Proof. Let the assertion be false, i.e. assume that S has no critical lattice. Denote by ϵ an arbitrarily small positive number, and by ρ so large a positive number that S is contained in the sphere

$$|X| < \rho.$$

---

[2]A sequence of lattices $\Lambda_1, \Lambda_2, \Lambda_3, \ldots$ is said to be *bounded* if (i) the determinants $d(\Lambda_r)$ are bounded, and (ii) no point $P \neq O$ of these lattices lies in a certain neighbourhood of O. (LP, Definition 1, p. 155.)

[3]It is possible to select from any bounded sequence of lattices a subsequence tending to a limiting lattice. (LP, Theorem 2, p. 156.)

Choose further any infinite sequence of S-admissible lattices

$$\Lambda_1, \Lambda_2, \Lambda_3, \ldots$$

satisfying (2). By Theorem 1, this sequence cannot be bounded. Hence there is an index k such that Λ_k contains a point $P_1 \neq O$ at a distance less than ϵ from O. There is no loss of generality in assuming that P_1 is of the form

$$P_1 = (\xi_1, 0, \ldots, 0), \text{ where } 0 < \xi_1 < \epsilon,$$

since the coordinate system may be so selected that the x_1-axis passes through P_1. Let now $P_2, P_3, \ldots, P_n$ be the points

$$P_2 = (0, \rho, 0, \ldots, 0), \quad P_3 = (0, 0, \rho, \ldots, 0), \ldots, P_n = (0, 0, 0, \ldots, \rho),$$

and let Λ be the lattice of basis $P_1, P_2, \ldots, P_n$, hence of determinant

$$d(\Lambda) = \xi_1\rho^{n-1} < \epsilon\rho^{n-1}. \tag{3}$$

Then this lattice is S-admissible. For Λ consists of the points

$$P = u_1P_1 + u_2P_2 + \ldots + u_nP_n \qquad (u_1, u_2, \ldots, u_n = 0, \mp 1, \mp 2, \ldots).$$

Of these lattice points, those with

$$\sum_{h=2}^{n} u_h^2 > 0$$

lie at a distance not less than ρ from 0, hence do not belong to S. If, however,

$$u_1 \neq 0, u_2 = u_3 = \ldots = u_n = 0,$$

then P belongs to Λ_k and so cannot be an inner point of S.

Hence

$$\Delta(S) \leq d(\Lambda) < \epsilon\rho^{n-1},$$

whence

$$\Delta(S) = 0$$

since ϵ may be arbitrarily small. Therefore S is of the zero type, contrary to hypothesis.

Theorem 2 contains as a special case my earlier result on the critical lattices of a star body of the finite type.[4]

II

4. The question arises whether Theorem 1 has a non-trivial content, thus whether there do in fact exist point sets satisfying the condition (1), but having no critical lattices. We shall now answer this problem by constructing an example of such a point set. But it will first be necessary to prove a number of simple lemmas.

5. Let

$$a_1, a_2, a_3, \ldots$$

be an infinite sequence of positive numbers satisfying

$$a_1 < a_2 < a_3 < \ldots, \quad \lim_{r\to\infty} a_r = \infty,$$

---

[4] LP, Theorem 8, p. 159.

and such that

$$\frac{\alpha_r}{\alpha_s} \text{ is irrational if } r \neq s.$$

Denote by Σ the set of all products

$$u\alpha_r, \text{ where } r, u = 1, 2, 3, \ldots .$$

Then all elements of Σ are positive; no two elements of Σ are equal; and any finite interval contains at most a finite number of elements of Σ. Hence if the elements of Σ,

$$\xi_1, \xi_2, \xi_3, \ldots \qquad \text{say,}$$

are arranged according to increasing size,

$$\xi_1 < \xi_2 < \xi_3 < \ldots,$$

then

$$\lim_{\mu \to \infty} \xi_\mu = \infty.$$

If t is any positive number, and if ξ_μ, ξ_ν run over all pairs of elements of Σ for which

$$\xi_\mu \neq \xi_\nu \text{ (i.e. } \mu \neq \nu), \xi_\nu \leq t,$$

then at most a finite number of the differences

$$| \xi_\mu - \xi_\nu |$$

are less than an arbitrary given constant. Denote by

$$\rho(t) = \min (| \xi_\mu - \xi_\nu |)$$

the smallest of these differences; it clearly defines a positive and non-increasing function of t.

Moreover,

$$\lim_{t \to \infty} \rho(t) = 0.$$

For Σ contains the elements,

$$u\alpha_1, v\alpha_2 \qquad (u, v = 1, 2, 3, \ldots),$$

and, as is well known, there are positive integers u, v, for which

$$| u\alpha_1 - v\alpha_2 |$$

is arbitrarily small.

6. From the definition of $\rho(t)$,

$$| \xi_\mu - \xi_\nu | \geq \max (\rho(\xi_\mu), \rho(\xi_\nu)), \qquad \text{if } \mu \neq \nu. \tag{4}$$

This implies that for no real number x both

$$| x - \xi_\mu | \leq \tfrac{1}{3}\rho(\xi_\mu) \quad \text{and} \quad | x - \xi_\nu | \leq \tfrac{1}{3}\rho(\xi_\nu),$$

unless $\mu = \nu$. For if, e.g. $\mu < \nu$, then from these inequalities,

$$| \xi_\mu - \xi_\nu | = | (x - \xi_\nu) - (x - \xi_\mu) | \leq \tfrac{1}{3}\rho(\xi_\mu) + \tfrac{1}{3}\rho(\xi_\nu) \leq \tfrac{2}{3}\rho(\xi_\mu) < \rho(\xi_\mu),$$

contrary to (4).

Lemma 1. *Let K be the set of all real numbers x satisfying at least one of the inequalities*

$$| x - \xi_\mu | \leq \tfrac{1}{3}\rho(2\xi_\mu) \qquad (\mu = 1, 2, 3, \ldots).$$

If all multiples

$$2^k x \qquad (k = 0, 1, 2, 3, \ldots)$$

of x belong to K, then x is an element of Σ.

Proof. From the hypothesis,

$$|2^k x - \xi_{\mu_k}| \le \tfrac{1}{6}\rho(2\xi_{\mu_k}) \qquad (k = 0, 1, 2, 3, \ldots),$$

where the indices μ_k depend on k. Therefore, in particular,

$$|2^{k+1}x - 2\xi_{\mu_k}| \le \tfrac{1}{3}\rho(2\xi_{\mu_k}),$$

$$|2^{k+1}x - \xi_{\mu_{k+1}}| \le \tfrac{1}{6}\rho(2\xi_{\mu_{k+1}}) < \tfrac{1}{3}\rho(\xi_{\mu_{k+1}}),$$

since $\rho(t)$ is a non-increasing function of t. But if ξ_μ belongs to Σ, so does $2\xi_\mu$; hence these inequalities imply that

$$\xi_{\mu_{k+1}} = 2\xi_{\mu_k} \qquad (k = 0, 1, 2, 3, \ldots),$$

whence

$$\xi_{\mu_k} = 2^k\xi_{\mu_0}, \; |2^k(x - \xi_{\mu_0})| \le \tfrac{1}{6}\rho(2^{k+1}\xi_{\mu_0}) \qquad (k = 0, 1, 2, 3, \ldots).$$

On letting k tend to infinity, the right-hand side tends to zero, and we find that

$$x = \xi_{\mu_0},$$

as asserted.

7. We need also the following, rather simpler, result.

LEMMA 2. *Let β be a positive number, and let K' be the set of all real numbers x satisfying at least one of the inequalities*

$$|x - u\beta| \le \frac{\beta}{6} \qquad (u = 1, 2, 3, \ldots).$$

If all multiples

$$2^k x \qquad (k = 0, 1, 2, 3, \ldots)$$

belong to K', then x is a positive integral multiple of β.

Proof. By hypothesis,

$$|2^k x - u_k\beta| \le \frac{\beta}{6} \qquad (k = 0, 1, 2, 3, \ldots)$$

with integers u_k depending on k. Therefore, in particular,

$$|2^{k+1}x - 2u_k\beta| \le \frac{\beta}{3},$$

$$|2^{k+1}x - u_{k+1}\beta| \le \frac{\beta}{6},$$

whence

$$|(u_{k+1} - 2u_k)\beta| = |(2^{k+1}x - 2u_k\beta) - (2^{k+1}x - u_{k+1}\beta)| \le \frac{\beta}{3} + \frac{\beta}{6} = \frac{\beta}{2},$$

and therefore

$$|u_{k+1} - 2u_k| \le \tfrac{1}{2},\; u_{k+1} = 2u_k,\; u_k = 2^k u_0 \qquad (k = 0, 1, 2, 3, \ldots),$$

since the u's are integers. Hence,

$$|2^k(x - u_0\beta)| \le \frac{\beta}{6} \qquad (k = 0, 1, 2, 3, \ldots).$$

On allowing k to tend to infinity, we find that

$$x = u_0\beta,$$

as asserted.

8. From the last two lemmas, we deduce a similar result for a special point set in n-dimensional space R_n.

Denote by

$$\alpha_1, \alpha_2, \alpha_3, \dots \quad \text{and} \quad \beta_1, \beta_2, \beta_3, \dots$$

two infinite sequences of positive numbers satisfying the following conditions:

(I)
$$\begin{aligned} &\alpha_1 < \alpha_2 < \alpha_3 < \dots, && \lim_{r\to\infty} \alpha_r = \infty, \\ &\beta_1 > \beta_2 > \beta_3 > \dots, && \lim_{r\to\infty} \beta_r = 0, \\ &\alpha_1\beta_1 > \alpha_2\beta_2 > \alpha_3\beta_3 > \dots, && \lim_{r\to\infty} \alpha_r\beta_r = 1. \end{aligned}$$

(II) If

$$\gamma_1, \gamma_2, \dots, \gamma_r$$

is any finite system of integers not all zero, then[5]

$$\alpha_1\gamma_1 + \alpha_2\gamma_2 + \dots + \alpha_r\gamma_r \neq 0.$$

Let further $u_1, u_2, \dots, u_n$ and r run over all positive integers, and denote by

$$\Pi^{(r)}(u) = \Pi^{(r)}(u_1, u_2, \dots, u_n)$$

the parallelepiped of all points

$$X = (x_1, x_2, \dots, x_n)$$

which satisfy the inequalities

$$|x_1 - \alpha_r u_1| \leq \tfrac{1}{6}\rho(2\alpha_r u_1), \ |x_2 - \beta_r u_2| \leq \frac{\beta_r}{6}, \ |x_h - u_h| \leq \tfrac{1}{6} \quad (h = 3, 4, \dots, n);$$

here $\rho(t)$ is the function defined in **5**. The centre of $\Pi^{(r)}(u)$ is at the point,

$$P^{(r)}(u) = P^{(r)}(u_1, u_2, \dots, u_n) = (\alpha_r u_1, \beta_r u_2, u_3, \dots, u_n).$$

Denote then by

$$\Pi = \bigcup_{u,r} \Pi^{(r)}(u)$$

the sum set of all parallelepipeds $\Pi^{(r)}(u)$, and by

$$\mathrm{P} = \{P^{(r)}(u)\}$$

the set of all points $P^{(r)}(u)$. Since, from (4),

$$\rho(2\xi_r) \leq \xi_r,$$

because both ξ_r and $2\xi_r$ belong to Σ, the two point sets Π and P lie completely in the octant

$$x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0.$$

---

[5]The conditions (I) and (II) are satisfied if, e.g.

$$\alpha_r = \left(1 + \frac{1}{r}\right)e^r, \quad \beta_r = \left(1 + \frac{1}{r}\right)e^{-r} \qquad (r = 1, 2, 3, \dots),$$

as is trivial for (I), and follows for (II) from the transcendency of e.

LEMMA 3. *Let the point* $X = (x_1, x_2, \ldots, x_n)$ *be such that all multiples*

$$2^k X = (2^k x_1, 2^k x_2, \ldots, 2^k x_n) \qquad (k = 0, 1, 2, 3, \ldots)$$

belong to Π. *Then* X *is an element of* P.

Proof. The first coordinate x_1 of X lies in one of the intervals

$$|x_1 - \xi_\nu| \leq \tfrac{1}{6}\rho(2\xi_\nu) \qquad (\nu = 1, 2, 3, \ldots);$$

the second coordinate x_2 lies in one of the intervals

$$|x_2 - \beta_r u_2| \leq \frac{\beta_r}{6} \qquad (r, u_2 = 1, 2, 3, \ldots);$$

and the remaining coordinates x_h $(h = 3, 4, \ldots, n)$ lie in intervals

$$|x_h - u_h| \leq \tfrac{1}{6} \qquad (u_h = 1, 2, 3, \ldots);$$

moreover, analogous conditions are also satisfied by the coordinates of the points

$$2^k X \qquad (k = 1, 2, 3, \ldots).$$

Therefore, by Lemma 1, x_1 belongs to Σ, so that

$$x_1 = a_r u_1 \tag{5}$$

for some pair of positive integers r and u_1. The same index r occurs in the inequalities for the multiples $2^k x_2$ of x_2; by Lemma 2 applied with $\beta = \beta_r$, there is therefore a positive integer u_2 such that

$$x_2 = \beta_r u_2 . \tag{6}$$

Finally, by the same lemma applied with $\beta = 1$, there exist $n-2$ positive integers $u_3, u_4, \ldots, u_n$ such that

$$x_h = u_h \qquad (h = 3, 4, \ldots, n). \tag{7}$$

The assertion is contained in (5), (6), and (7).

9. We also need the following simple lemma about the bases of a lattice.

LEMMA 4. *For every lattice* Λ, *a basis*

$$Y_1 = (y_{11}, y_{12}, \ldots, y_{1n}),\ Y_2 = (y_{21}, y_{22}, \ldots, y_{2n}), \ldots,\ Y_n = (y_{n1}, y_{n2}, \ldots, y_{nn})$$

can be found such that

$$y_{hk} > 1 \qquad (h, k = 1, 2, \ldots, n). \tag{8}$$

Proof. First choose an arbitrary point $Y_1 = (y_{11}, y_{12}, \ldots, y_{1n})$ of Λ with

$$y_{11} > 1,\ y_{12} > 1, \ldots, y_{1n} > 1$$

such that no inner point of the line segment joining O to Y_1 belongs to Λ. By Minkowski's selection method[6], $n-1$ further lattice points $Y'_2, Y'_3, \ldots, Y'_n$ can be chosen such that the n points

$$Y_1, Y'_2, Y'_3, \ldots, Y'_n$$

form a basis of Λ. Then the further n points

$$Y_1,\ Y_2 = Y'_2 + v_2 Y_1,\quad Y_3 = Y'_3 + v_3 Y_1, \ldots,\quad Y_n = Y'_n + v_n Y_1$$

where $v_2, v_3, \ldots, v_n$ are $n-1$ arbitrary integers, also form a basis of Λ. We satisfy now the conditions (8) by taking the v's positive and sufficiently large.

---

[6] *Geometrie der Zahlen*, § 46.

10. As in **8**, we let $u_1, u_2, \ldots, u_n$ and r run over all positive integers, but denote now by

$$\Pi_0^{(r)}(u) = \Pi_0^{(r)}(u_1, u_2, \ldots, u_n)$$

the *open* parallelepiped of all points X satisfying

$$|x_1 - \alpha_r u_1| < \tfrac{1}{6}\rho(2\alpha_r u_1), |x_2 - \beta_r u_2| < \frac{\beta_r}{6}, |x_h - u_h| < \tfrac{1}{6} \; (h = 3, 4, \ldots, n);$$

its centre is again at the point $P^{(r)}(u)$, and its closure is $\Pi^{(r)}(u)$. Further denote by

$$\Pi_0 = \bigcup_{u,r} \Pi_0^{(r)}(u)$$

the sum of all parallelepipeds $\Pi_0^{(r)}(u)$, and by Ω the point set

$$x_1 \geq 1, x_2 \geq 1, \ldots, x_n \geq 1.$$

The difference set

$$S = \Omega - \Pi_0$$

of all points of Ω which are not in Π_0, is evidently closed, since Π_0, as a sum of open sets, is open, and since Ω is closed because R_n does not contain a point at infinity.

There are at most a finite number of points $P^{(r)}(u)$ in every finite portion of Ω. Therefore every point of S is either an inner point of S, or a boundary point of Ω, or it is a boundary point of one of the closed parallelepipeds $\Pi^{(r)}(u)$, hence belongs to Π.

11. Let now Λ be any S-admissible lattice. Then choose a basis $Y_1, Y_2, \ldots, Y_n$ of Λ satisfying the condition (8) of Lemma 4. These n points, and also the vector sum

$$Y = Y_1 + Y_2 + \ldots + Y_n$$

are not inner points of S, nor are they boundary points of Ω; and the same is true even for the multiples

$$(9) \quad 2^k Y_1, 2^k Y_2, \ldots, 2^k Y_n, 2^k(Y_1 + Y_2 + \ldots + Y_n) = 2^k Y \; (k = 0, 1, 2, 3, \ldots).$$

Hence all points (9) belong to Π. But then, by Lemma 3, the $n + 1$ points

$$Y_1, Y_2, \ldots, Y_n, Y$$

are elements of P, and so there exist positive integers

$$r_1, r_2, \ldots, r_n, r$$

and

$$u_{hk}, u_k \qquad (h, k = 1, 2, \ldots, n)$$

such that

$$Y_h = (\alpha_{r_h} u_{h1}, \beta_{r_h} u_{h2}, u_{h3}, \ldots, u_{hn}) \qquad (h = 1, 2, \ldots, n),$$

$$Y = Y_1 + Y_2 + \ldots + Y_n = (\alpha_r u_1, \beta_r u_2, u_3, \ldots, u_n).$$

Therefore, in particular,

$$\alpha_{r_1} u_{11} + \alpha_{r_2} u_{21} + \ldots + \alpha_{r_n} u_{n1} = \alpha_r u_1.$$

By the hypothesis (II) of **8**, this equation can hold only if

$$r_1 = r_2 = \ldots = r_n = r, \quad u_{11} + u_{21} + \ldots + u_{n1} = u_1.$$

Hence all basis points Y_h belong to the same value of r, and the basis is of the form

$$Y_h = (\alpha_r u_{h1}, \beta_r u_{h2}, u_{h3}, \ldots, u_{hn}) \qquad (h = 1, 2, \ldots, n).$$

Denote now by Λ_r the lattice of all points

$$P = (\alpha_r g_1, \beta_r g_2, g_3, \ldots, g_n),$$

where the g's run over all integers; this lattice is of determinant

$$d(\Lambda_r) = \alpha_r \beta_r.$$

Since the basis elements Y_h of Λ belong to Λ_r, Λ is either identical with Λ_r or it is a sublattice. In either case,

$$d(\Lambda) = g d(\Lambda_r),$$

where g is a positive integer. Hence, by the hypothesis (I) of **8**,

$$d(\Lambda) \geq d(\Lambda_r) > 1, \quad \text{and } d(\Lambda) > 2 \text{ if } g > 1.$$

In the other direction, from the same hypothesis,

$$\lim_{r \to \infty} d(\Lambda_r) = 1.$$

We find therefore the following result:

THEOREM 4. *The only admissible lattices of the set S are* (i) *the lattices* $\Lambda_1, \Lambda_2, \Lambda_3, \ldots,$ *and* (ii) *their sublattices. All S-admissible lattices are of determinant greater than* 1, *but*

$$\lim_{r \to \infty} d(\Lambda_r) = 1.$$

Hence $\Delta(S) = 1$, and there are no critical lattices of S.

12. Theorem 4 implies, in particular, that S has only an ***enumerable*** set of admissible lattices, a possibility which cannot arise for star bodies. It is further clear that no point of any S-admissible lattice lies on the boundary of S.

The following, somewhat simpler, example of a point set is possibly even more surprising. Denote by T the set of all points X such that

$$\max(|x_1 - u_1|, |x_2 - u_2|, \ldots, |x_n - u_n|) \geq \tfrac{1}{8}$$

for every system of integers $u_1, u_2, \ldots, u_n$. It is not difficult to deduce from Lemma 2, that the only T-admissible lattices are (i) the lattice of all points with integral coordinates, and (ii) all its sublattices. Therefore $\Delta(T) = 1$, and there is just one critical lattice. Every point of this critical lattice lies at a distance $\frac{1}{8}$ from the boundary of T, and the same is true for the points of the T-admissible lattices. This is very different from the position for

star bodies; for every critical lattice of a star body has at least one point arbitrarily near to its boundary.

POSTSCRIPT (*June* 1948)

Mr. C. A. Rogers, having been told of my result, found the following simpler example of a point set without a critical lattice:

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_1x_2\left(1 - \frac{x_1x_2}{x_1^2 + x_2^2}\right) \leq 1 .$$

This two-dimensional set differs from my example in having a *continuous* infinity of admissible lattices.

University of Manchester

THE NONEXISTENCE OF CERTAIN FINITE PROJECTIVE PLANES

R. H. BRUCK and H. J. RYSER

1. Introduction. A projective plane geometry π is a mathematical system composed of undefined elements called points and undefined sets of points (at least two in number) called lines, subject to the following three postulates:

(P_1) Two distinct points are contained in a unique line.

(P_2) Two distinct lines contain a unique common point.

(P_3) Each line contains at least three points.

The projective plane π is *finite* if it consists of a finite number of points. If π is finite, then there exists a positive integer N such that each line of π contains exactly $N + 1$ distinct points, and each point is contained in exactly $N + 1$ distinct lines. Moreover, π has exactly $N^2 + N + 1$ distinct points and $N^2 + N + 1$ distinct lines (see [3], [6], [13]).

In all known finite geometries the integer N is a power of a prime. Indeed, for every prime p and for every positive integer n, finite geometries with $N = p^n$ have been constructed by means of the Galois fields $GF[p^n]$ (see [12]). It is still an unsettled question whether or not N must be the power of a prime. In this connection it has been shown that there does not exist a finite geometry for $N = 6$ (see [11]). The purpose of our paper is to prove the following more general theorem on the non-existence of finite geometries.

THEOREM 1. *If $N \equiv 1$ or $2 \bmod 4$ and if the square free part of N contains at least one prime factor of the form $4k + 3$, then there does not exist a finite projective plane geometry with $N + 1$ points on a line.*

In section 2 finite geometries are studied in connection with matrices whose elements are non-negative integers. The Minkowski-Hasse theory on the equivalence of quadratic forms under rational transformations is discussed in section 3, and the results of sections 2 and 3 are then utilized in section 4 to prove Theorem 1.

It is to be noted that Theorem 1 asserts in particular that a geometry does not exist for $N = 2p$, where p is a prime of the form $4k + 3$. Moreover, a finite plane with $N + 1$ points on a line can always be constructed from a given complete set of mutually orthogonal Latin squares of order $N \geqq 3$ (see [1], [8]). Thus for any N of Theorem 1 there does not exist a complete set of mutually orthogonal Latin squares of order N.

2. The Incidence Matrix. An n-rowed square matrix A each of whose elements is zero or one is an *incidence matrix* provided it satisfies the following three conditions:

Received May 7, 1948.

(I_1) If r_1 and r_2 are two distinct rows of A, then there is a unique integer j such that the rows r_1 and r_2 each have the integer one in the jth column.

(I_2) If c_1 and c_2 are two distinct columns of A, then there is a unique integer i such that the columns c_1 and c_2 each have the integer one in the ith row.

(I_3) Each row of A contains at least three ones.

THEOREM 2. *If π is a finite projective plane geometry with $N + 1$ points on a line, then there exists an incidence matrix A of order $n = N^2 + N + 1$. If A^T denotes the transpose of the matrix A, then*

$$B = AA^T = A^TA, \tag{M}$$

where B is an integral matrix with $N + 1$ down the main diagonal and ones in all other positions.

For let the $N^2 + N + 1$ points of π be numbered in any convenient order $1, 2, \ldots, N^2 + N + 1$ and listed in a row. Let the $N^2 + N + 1$ lines be numbered similarly $1, 2, \ldots, N^2 + N + 1$ and listed in a column. Then let a table of $N^2 + N + 1$ rows and $N^2 + N + 1$ columns be formed by inserting a one in row i and column j if line i contains point j, and a zero in the contrary case. Then by the properties of the geometry π given in section 1, it follows that the table yields an incidence matrix A which satisfies the equation (M).

THEOREM 3. *If a matrix A with non-negative integral elements and of order $n > 1$ satisfies the equation* (M), *where $N \geqq 2$, then A is an incidence matrix and defines a finite projective plane geometry with $N + 1$ points on a line.*

The matrix A must be composed entirely of zeros and ones. For if a_{ij} were an element of A in row i and column j and if a_{ij} were greater than one, then by equation (M) each element in column j of A except a_{ij} would be zero. Moreover, each element in row i of A except a_{ij} would also be zero. But then the matrix AA^T would contain a zero element, and this is impossible if A is to satisfy (M). Since A is composed of zeros and ones and since A satisfies (M) with $N \geqq 2$, it follows that A is an incidence matrix, and this incidence matrix can be used to define the finite projective plane.

3. Congruence of Matrices. Let A and B be two symmetric matrices of order n with elements in the rational field. The matrices A and B are *congruent*, written $A \sim B$, provided there exists a non-singular matrix C with rational elements such that

$$A = C^TBC.$$

It is easy to show that congruence of matrices satisfies the usual requirements of an equals relationship.

non-singular

Suppose now that A is an integral symmetric matrix of order and rank n. It is well known that one can always construct an integral diagonal matrix $D = [d_1, d_2, \ldots, d_n]$, where $d_i \neq 0$ for $i = 1, 2, \ldots, n$, such that $D \sim A$.

The number of negative terms ι in this diagonal is called the *index* of A. Sylvester's law of inertia states that ι is an invariant of A (see [7]).

Let $d = (-1)^{\iota}\delta$, where δ is the square free positive part of the determinant $|A|$ of the matrix A. From the matric equation $B = C^{T}AC$, it follows that $|B| = |C|^2|A|$. Hence d is a second invariant of A.

Minkowski [9] and Hasse [4] have introduced a third invariant c_p, which with the preceding two completes the system. Before discussing the invariant c_p, we recall now the essentials of the Hilbert norm-residue symbol $(m,n)_p$. The norm-residue symbol is defined for arbitrary non-zero integers m and n and for every prime p. Its precise definition as well as complete proofs of the following two theorems can be found in the collected works of Hilbert [5].

THEOREM 4. *If m and n are integers not divisible by the odd prime p, then*

$$(m, n)_p = +1, \tag{1}$$

$$(n, p)_p = (p, n)_p = (n|p), \tag{2}$$

where $(n|p)$ is the Legendre symbol. Moreover, if $n \equiv m \not\equiv 0 \bmod p$, then

$$(m, p)_p = (n, p)_p. \tag{3}$$

THEOREM 5. *For arbitrary non-zero integers m, m', n, n' and for every prime p,*

$$(-n, n)_p = +1, \tag{4}$$

$$(m, n)_p = (n, m)_p, \tag{5}$$

$$(mm', n)_p = (m, n)_p(m', n)_p, \tag{6}$$

$$(n, mm')_p = (n, m)_p(n,m')p. \tag{7}$$

At this point it is convenient to prove a Lemma which is useful for the proof of Theorem 1 in section 4.

LEMMA. *For p an odd prime and for every positive integer n,*

$$(n, n+1)_p = (-1, n+1)_p, \tag{8}$$

$$(n, n^2+n+1)_p = +1, \tag{9}$$

$$\prod_{i=1}^{n} (i, i+1)_p = ((n+1)!, -1)_p. \tag{10}$$

If p does not divide n or $n+1$, then (8) is trivial. If p divides n, then $n+1 \equiv 1 \bmod p$ and if p divides $n+1$, then $n \equiv -1 \bmod p$. By (3) of Theorem 4 equation (8) is established. If p divides n, then $n^2+n+1 \equiv (n+1)^2 \not\equiv 0 \bmod p$ and if p divides n^2+n+1, then $n \equiv (n+1)^2 \not\equiv 0 \bmod p$. This establishes (9). Equation (10) is a consequence of (8) and Theorem 5.

Now let A be a non-singular and symmetric integral matrix of order n. Let D_r denote the leading principal minor determinant of order r, and suppose that $D_r \neq 0$ for $r = 1, 2, \ldots, n$. The invariant c_p is then defined for every odd prime p by the equation

$$c_p = c_p(A) = (-1, -D_n)_p \prod_{i=1}^{n-1} (D_i, -D_{i+1})_p.$$

By (1) of Theorem 4, evidently $c_p = -1$ for only a finite number of p.

We are now in a position to state the fundamental Minkowski-Hasse theorem, a proof of which can be found in the original paper of Hasse [4]. More recent developments of the theory are discussed in [2] and [10].

THEOREM 6. *Let A and B be two integral symmetric matrices of order and rank n. Suppose further that the leading principal minor determinants of A and B are different from zero. Then $A \sim B$ if and only if A and B have the same invariants ι, d, and c_p for every odd prime p.*

4. Proof of Theorem 1. Let N be a positive integer and let B_n denote the integral matrix of order n with $N+1$ down the main diagonal and ones in all other positions. If we subtract column one of B_n from each of the other columns, and then add to row one each of the other rows, we obtain

$$|B_n| = N^{n-1}(N+n).$$

In particular if $n = N^2 + N + 1$, then B_n is the matrix B of equation (M) and $|B|$ is the square of an integer.

If row n of B_n is subtracted from each of the other rows, and if column n is then subtracted from each of the other columns, the resulting matrix is

$$Q_n = \begin{bmatrix} 2N & N & N & \dots & -N \\ N & 2N & N & \dots & -N \\ N & N & 2N & \dots & -N \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ -N & -N & -N & \dots & N+1 \end{bmatrix},$$

and this matrix is congruent to B_n. Hence for every odd prime p, $c_p(B_n) = c_p(Q_n)$. Moreover, if E_i denotes the determinant of order i with $2N$ down the main diagonal and N in all other positions, then $E_i = N^i(i+1)$. Thus if $n = N^2 + N + 1$ and if p is an odd prime, then the invariant $c_p(B) = c_p(Q_n)$ of the matrix B of equation (M) is given by

$$c_p(B) = (E_{n-1}, -1)_p \prod_{i=1}^{n-2} (E_i, -E_{i+1})_p.$$

In the subsequent computation we prove

$$c_p(B) = (-1, N)_p^{\frac{N(N+1)}{2}}. \tag{E}$$

By Theorem 5 and (10), and omitting for convenience the subscript p,

$$\prod_{i=1}^{n-2} (E_i, -E_{i+1}) = \prod_{i=1}^{n-2} (N^i(i+1), -N^{i+1}(i+2))$$

$$= \prod_{i=1}^{n-2} (N^i, -N^{i+1})\,(i+1, -(i+2))\,S$$

$$= (N, -1)^{\frac{(n-1)(n-2)}{2}} ((n-1)!, -1)\,(n!, -1)\,S,$$

where

$$S = \prod_{i=1}^{n-2} (N^i, i+2)\,(N^{i+1}, i+1).$$

Moreover, by (9)

$$S = \prod_{i=1}^{n-2} (N^i, i+2) \prod_{i=0}^{n-3} (N^i, i+2)$$
$$= (N,n)^{n-2} = +1.$$

Thus

$$c_p(B) = (N^{n-1}n, -1)\,(N, -1)^{\frac{(n-1)(n-2)}{2}}\,(n, -1)$$
$$= (N, -1)^{n-1}(N, -1)^{\frac{(n-1)(n-2)}{2}} = (N, -1)^{\frac{N(N+1)}{2}},$$

and this establishes equation (E).

Suppose now that π is a finite projective plane with $N+1$ points on a line. Then by equation (M) of section **2**, the matrix B is congruent to the identity matrix I. Since $c_p(I) = +1$ for every odd prime p, it follows that if π exists, then for every odd prime p,

$$c_p(B) = (-1, N)^{\frac{N(N+1)}{2}} = +1\,.$$

If now $N \equiv 1$ or $2 \bmod 4$, then the exponent $\dfrac{N(N+1)}{2}$ is odd. Moreover, if a prime p of the form $4k+3$ divides the square free part of N, then $(-1, N)_p = -1$. This is a contradiction and completes the proof of Theorem 1.

Postscript (*November* 13, 1948)

(*a*) In a letter to one of the authors, dated May 11, 1948, Marshall Hall pointed out that the n-rowed symmetric matrix B of section 4 ($n = N^2 + N + 1$) is the matrix of a quadratic form which can be written as

$$(x_2 + \ldots + x_n)^2 + N\left(x_2 + \frac{x_1}{N}\right)^2 + \ldots + N\left(x_n + \frac{x_1}{N}\right)^2.$$

Hall's remark demonstrates concretely that B is rationally congruent to the diagonal matrix $D = (1, N, N, \ldots, N)$ and thus permits a simpler derivation of equation (E).

(*b*) In 1782 Euler conjectured that a pair of orthogonal latin squares (or a graeco-latin square) of order N cannot exist if N has the form $4k+2$. The truth of Euler's conjecture would ensure (see [1], [8]) the non-existence of projective planes with $N \equiv 2 \bmod 4$ and hence would both imply and improve

one half of Theorem 1. For this reason the authors have decided to add to the bibliography a paper by H. F. MacNeish [14] containing a "proof" of Euler's conjecture. The correctness of this proof, however, has been questioned by F. W. Levi. In this connection see [6] (Second Lecture); *Jahrbuch der Math.*, vol. 48 (1921), 71; *Jahrbuch der Math.*, vol. 49 (1923), 41–42.

References

[1] R. C. Bose, "On the application of the properties of Galois fields to the problem of construction of hyper-Graeco-Latin squares," Sankhya, *Indian Journal of Statistics*, vol. 3 (1938), 323-338.

[2] W. H. Durfee, "Quadratic forms over fields with a valuation," *Bull. Amer. Math. Soc.*, vol. 54 (1948), 338-351.

[3] M. Hall, "Projective planes," *Trans. Amer. Math. Soc.*, vol. 54 (1943), 229-277.

[4] H. Hasse, "Über die Äquivalenz quadratischer Formen im Körper der rationalen Zahlen," *J. reine angew. Math.*, vol. 152 (1923), 205-224.

[5] D. Hilbert, *Gesammelte Abhandlungen*, I (Berlin, 1932), 161-173.

[6] F. W. Levi, *Finite geometrical systems* (University of Calcutta, 1942).

[7] C. C. MacDuffee, *The theory of matrices* (New York, 1946), 56.

[8] H. B. Mann, "On orthogonal Latin squares," *Bull. Amer. Math. Soc.*, vol. 50 (1944), 249-257.

[9] H. Minkowski, *Gesammelte Abhandlungen*, I (Leipzig and Berlin, 1911), 219-239.

[10] G. Pall, "The arithmetical invariants of quadratic forms," *Bull. Amer. Math. Soc.*, vol. 51 (1945), 185-197.

[11] G. Tarry, "Le problème de 36 officiers," *Compte Rendu de l'Association Française pour l'Avancement de Science Naturel*, vol. 1 (1900), 122-123, vol. 2 (1901), 170-203.

[12] O. Veblen and W. H. Bussey, "Finite projective geometries," *Trans. Amer. Math. Soc.*, vol. 7 (1906), 241-259.

[13] O. Veblen and J. H. M. Wedderburn, "Non-Desarguesian and non-Pascalian geometries," *Trans. Amer. Math. Soc.*, vol. 8 (1907), 379-388.

[14] H. F. MacNeish, "Euler squares," Ann. of Math., vol. 23 (1921-22), 221-227.

The University of Wisconsin

GENERALIZED VECTOR SPACES. I.

The Structure of Finite-Dimensional Spaces

KARL MENGER

1. Introduction

During the last fifty years, the concept of the Euclidean space (an n-dimensional coordinate space with a Pythagorean distance) has undergone various profound generalizations.

Hilbert introduced the infinitely-dimensional Euclidean space whose points are infinite sequences of coordinates having from the origin, and thus from each other, finite Pythagorean distances.

Minkowski generalized the Pythagorean distance. Any surface which is symmetric about the origin, o, and intersects every ray issuing from o in exactly one point, is admitted as the "unit sphere" about o, that is, as the set of all points having the distance 1 from o. The distance from o to a point whose coordinates are k times those of a point on the unit sphere, is k. Minkowski chose a congruent unit sphere about every point. He discovered the equivalence of the convexity of these spheres and the triangle inequality for the distance. Finsler introduced spaces which are locally Minkowskian in the same sense in which Riemann spaces are locally Euclidean. With each point, a "tangential" Minkowskian space is associated—the unit sphere varying from point to point. Finsler found that each positively definite problem of the Calculus of Variations gives rise to one of his spaces.

In the finite-dimensional case, Weyl noticed that the definition of points by coordinates could be replaced by the assumption that undefined points can be added, and multiplied by real numbers. Banach, Hahn, and Wiener [1], independently of each other, introduced the following concept. A set of elements, $v, w, \ldots$ (called vectors) is said to be a *vector space* if

(*a*) the set is a commutative group, the operation being denoted by $+$, the neutral element by o, so that $v + o = o + v = v$;

(*b*) an associative and doubly distributive multiplication of vectors by real numbers, $\alpha, \beta, \ldots$ is defined, that is to say,

$$\alpha(\beta v) = (\alpha\beta)v, \quad (\alpha + \beta)v = \alpha v + \beta v, \quad \alpha(v + w) = \alpha v + \alpha w;$$

for the multiplication by the numbers 1, -1, and 0 we have

$$1v = v, \quad -1v = -v, \quad 0v = o;$$

(*c*) with each vector, v, a real number $|v|$ is associated, called the norm of v, which satisfies the following three conditions

(1) $|\alpha v| = |\alpha||v|$ for every v;

(2) $|v + w| \leq |v| + |w|$;

(3) if $v \neq o$, then $|v| > 0$.

Received October 1, 1948.

Much earlier, Fréchet had introduced the most radical generalization of the Euclidean space by assuming only that a number (called distance) be associated with every unordered pair of elements of a set, identical elements having the distance 0, distinct elements a positive distance, while the distance satisfies the triangle inequality. As a price for the generality of these metric spaces we have to accept the possible absence of directions of any kind.

In applying metric methods to the Calculus of Variations we made use of all these generalizations of the concept of space [2]. We studied the minima of line integrals even in a general metric space. Our integrand is a function of the point and (in absence of a direction) of an ordered pair of distinct points. Multiplying the distance by this function we obtain a new distance which we call the *variational distance*. If, in particular, the metric space is a vector space, and the function is positive and endowed with strong continuity properties, then one obtains a Finsler space. If the metric space is Euclidean, then our results generalized Tonelli's existence theorems for the parametric case.

Besides synthesizing the various known concepts, the metric ideas in the Calculus of Variations led to a generalization of the idea of space in a new direction. Minkowski spaces as well as the vector spaces of Banach, Hahn, and Wiener, and even Fréchet's metric spaces, have the following two important features in common: distinct points have distances $\neq 0$; and distances are non-negative. But, on every level of generality, the only source of the lower semi-continuity of the line integral is the local triangle inequality of the variational distance. The two other traditional features (and still more, of course, the symmetry) of the distance appeared to be quite inessential. As a result, one can, in particular, generalize Finsler's concept in such a way that one can associate a generalized Finsler space also with semi-definite and indefinite parametric variational problems in vector spaces.

As a by-product, these studies yielded a generalization of the concept of a vector space. Alt [3] proved the equivalence of the triangle inequality with what he called "projective convexity" of the unit sphere. Pauc [4] and Aronszajn continued this work in many interesting ways and the latter first explicitly formulated the concept of general vector spaces [5] which implicitly was contained in our remarks [6] about what we called "generalized Minkowskian metric." In our spaces we had admitted that distinct points might have the distance 0, and that the distance of two points might be negative. In fact, we mentioned that vectors might have negative norms or the norm 0. We had dropped the symmetry of the distance and the norm. All we had retained was the triangle inequality for distances and norms, and the assumption that by multiplying a vector by a positive number, k, the norm was multiplied by k.

Now we intend to study these generalized vector spaces in a series of papers. The present first paper contains a few remarks about all generalized vector spaces but essentially deals with the structure of generalized vector spaces of finite dimension. We prove that each such space is built up of a subspace

all of whose vectors (except o) have a positive norm; a subspace all of whose vectors have the norm 0; and possibly one single line containing a vector with a non-positive norm while the norm of the opposite vector is positive.

In subsequent papers we shall study spaces of infinitely many dimensions, metric properties of our spaces as well as topological aspects of the theory ("triangular topologies"), non-real multipliers, and applications which, besides the Calculus of Variations, comprise the theories of operators and of normed rings.

2. The Main Types of Generalized Vector Spaces

A ***generalized vector space*** is a set, V, of elements for which addition, and multiplication by real numbers, are defined according to Postulates (a) and (b) while the norm $|v|$ of a vector is a real number satisfying only one and one half of the three Postulates (c), namely,

$$(1^+) \qquad \text{If } a > 0, \text{ then } |av| = a|v|.$$

$$(2) \qquad |v + w| \leq |v| + |w|.$$

We do not postulate the other half of (1), that is

$$(1^-) \qquad \text{If } a < 0, \text{ then } |av| = -a|v|,$$

nor the two important properties of the ordinary vector spaces which are jointly postulated in (3), namely, that $|v| \geq 0$ and that $v \neq o$ implies $|v| \neq 0$.

We shall briefly call a vector, v, ***positive***, ***negative*** or ***null*** according to whether $|v|$ is positive, negative or 0. We call the vector v ***degenerate*** if $v \neq o$ and $|v| = |-v| = 0$.

We call a general vector space, V,

definite if every vector, except o, is positive;
semi-definite if V is not definite but no vector is negative and at least one vector is positive;
indefinite if V contains both positive and negative vectors;
degenerate if V contains at least one degenerate vector;
non-degenerate if V is not degenerate;
totally degenerate if every vector of V is degenerate and thus null.

The vector spaces of Banach, Hahn, and Wiener are the definite vector spaces. If by O^* we mean the set containing only a vector o, then, according to the above terminology, O^* is a definite vector space, and consequently non-degenerate. In fact, O^* is a vector space in the ordinary sense.

That we use the terms definite and semi-definite instead of ***positively*** definite and ***positively*** semi-definite will not lead to ambiguities since we shall see in Section 4 that no space is negatively definite or negatively semi-definite. There would be only negatively definite and negatively semi-definite spaces if we postulated 1^+) in conjunction with the triangle contra-inequality

$$|v + w| \geq |v| + |w|.$$

3. Vector-Algebraic Preliminaries

A subset, V', of V is called a ***subspace*** if for every two vectors, v and w, of V' and for every number a, the vectors $v + w$ and av belong to V'. The set consisting of o alone is the subspace O^*.

If S is a subset of V, then we denote by $V(S)$ the subspace of V consisting of all vectors $a_1v_1 + a_2v_2 + \ldots + a_nv_n$ where n is any integer, the a_i are numbers, the v_i vectors of S. In particular, if S consists of only one vector $v \neq o$, then $V(S)$ is called the ***v-line*** and denoted by $[v]$. In the usual way, we mean by $V' + V''$, the ***join*** of two subspaces V' and V'', the subspace $V(S)$ where S is the set of all vectors belonging to V' and/or V''; by $V' . V''$, the ***intersection*** of V' and V'', the subspace of all vectors belonging to both V' and V''. If $V' \neq O^* \neq V''$ and $V' . V'' = O^*$, then V' and V'' are called ***independent***.

Lemma. *If V' is a subspace of V, then there exists a subspace, V'', of V such that $V' + V'' = V$ and $V' . V'' = O^*$.*

If $V' = V$, then $V'' = O^*$. If $V' \neq V$, then there exists a vector, v_1, which does not belong to V'. In this case, let Ω be any ordinal number about which we make the following assumption: with every ordinal number $\omega < \Omega$ a vector, v_ω, has been associated in such a way that if S_ω is the set of all vectors $v_1, \ldots, v_\omega$, then

(1) the set S_ω does not contain any finite subset of dependent vectors;
(2) $V(S_\omega) . V' = O^*$.

We call T_Ω the set of all vectors v_ω such that $\omega < \Omega$. Then two cases are possible. Either $V = V(T_\Omega)$ in which case we set $V'' = V(T_\Omega)$ and our proposition holds. Or V contains vectors not belonging to the join $V' + V(T_\Omega)$. In this case, we call one of these vectors v_Ω, and denote the set of all vectors $v_1, \ldots, v_\Omega$ by S_Ω. Then we have associated a vector v_ω with every ordinal number $\omega \leq \Omega$ in such a way that conditions (1) and (2) are satisfied. There exists an ordinal number Ω such that the first case prevails. If V is n-dimensional, this follows by induction, and $\Omega \leq n$. If V is infinitely dimensional, the conclusion is valid by transfinite induction.

If $v \neq o$, then we call the set of the vectors av for all $a > 0$, the open v-ray or, briefly, since we shall not consider rays which include o, the *v-ray*. We call the $(-v)$-ray the ***opposite*** ray. The v-line consists of o, the v-ray, and the opposite ray.

If v and w are independent vectors (that is, vectors $\neq o$ neither lying on the line of the other), then we call the set of the vectors $av + \beta w$ for all real numbers a, β, the *v, w-plane*. We further call the set of the vectors $av + \beta w$ such that $a \geq 0$, $\beta \geq 0$ ($a > 0$, $\beta > 0$) the *closed* (*open*) *v, w-quadrant*. We denote these quadrants by $[v, w]$ and (v, w), respectively. We can also introduce the half-open quadrants $[v, w)$ and $(v, w]$.

The set of all vectors which are opposite to those of the open, the closed, the half-open first v, w-quadrants are called the open, the closed, the half-open ***third v, w-quadrants***, respectively. We denote these sets by $)v, w($, $]v,w[$,

$]v,w($, $)v,w[$, respectively. Clearly, the first and third quadrants are associated with the unordered vector pair, v, w. With the ordered pair v, w we can also associate the closed *second* v, *w-quadrant*, that is, the set $]v,w]$ of the vectors $-\alpha v + \beta w$ such that $\alpha \geq 0$, $\beta \geq 0$. Similarly we define $)v,w)$, $]v,w)$, $)v,w]$, and the *fourth v,w-quadrants*. One readily proves

REMARK 1. *If v, w, x are pairwise independent vectors and x belongs to (v,w), then a vector y belongs to (v,w) if and only if y either belongs to (v,x) or to (x,w) or to the x-ray.*

REMARK 2. *If v,w,x are pairwise independent and x belongs to $)v,w($, then every vector of the v,w-plane belongs to (v,w) or to (v,x) or to (w,x) or to the rays of one of the vectors v,w,x.*

4. COROLLARIES OF THE ASSUMPTIONS ABOUT THE NORM

We shall deduce immediate consequences of the assumptions (1^+) and (2) about the norm in a generalized vector space.

If in (1^+) we set $\alpha = 2$ and $v = o$, then since $2o = o$ we conclude

$$(1^\circ) \qquad |o| = 0.$$

If in (2) we set $v = v' + v''$ and $w = v''$, we obtain

$$|v'| = |v' + v'' - v''| \leq |v' + v''| + |-v''|,$$

thus

$$|v' + v''| \geq |v'| - |-v''|.$$

Similarly, $|v' + v''| \geq |v''| - |-v'|$. Hence

$$(2a) \qquad \text{Max}\,[\,|v| - |-w|, |w| - |-v|\,] \leq |v + w| \leq |v| + |w|.$$

If in (2a) we set $w = -v$, then by (1°) we have $0 \leq |v| + |-v|$ and thus

$$|v| \geq -|-v| \quad \text{and} \quad |-v| \geq -|v|.$$

In particular, we can formulate the following

LEMMA. *The opposite of a negative vector is positive. The opposite of a null vector is non-negative.*

As a corollary of this lemma we see that *no space is negatively definite or negatively semi-definite.*

In absence of a general concept of limit, we can prove only two restricted continuity properties of the norm.

ADDITIVE CONTINUITY. *For every $\epsilon > 0$ there exists a $\delta > 0$, namely $\delta = \epsilon$, such that for every vector v*

from $|w| < \delta$ and $|-w| < \delta$ it follows that $|v| - \delta < |v + w| < |v| + \delta$.

This is an immediate consequence of 2a).

FINITE-DIMENSIONAL CONTINUITY. *For every $\epsilon > 0$ every integer n, and every n-tuple of vectors $w_1, w_2, \ldots, w_n$, there exists a $\delta > 0$ (depending upon $\epsilon, w_1, \ldots, w_n$) such that from*

$$|\delta_1| < \delta, |\delta_2| < \delta, \ldots, |\delta_n| < \delta$$

for every vector v it follows that

$$|v| - \epsilon \leq |v + \delta_1 w_1 + \delta_2 w_2 + \ldots + \delta_n w_n| \leq |v| + \epsilon.$$

Setting $\delta_1 w_1 + \ldots + \delta_n w_n = w$ we see that both $|w|$ and $|-w|$ are $\leq n \operatorname{Max} |\delta_i| \cdot \operatorname{Max} [\,|w_i|, |-w_i|\,]$.

Thus $\delta = \frac{\epsilon}{n} \operatorname{Max} [\,|w_i|, |-w_i|\,]$ satisfies the requirement.

5. Lemmas

Lemma 1. *If v and w are independent non-positive vectors, then every vector of $[v, w]$, i.e., the closed first v, w-quadrant, is non-positive.*

For if $\alpha \geq 0$ and $\beta \geq 0$, then

$$|\alpha v + \beta w| \leq |\alpha v| + |\beta w| = \alpha |v| + \beta |w|.$$

The last expression is ≤ 0 if $|v| \leq 0$ and $|w| \leq 0$.

The last expression is < 0 if $\alpha > 0$, $\beta > 0$ and at least one of the vectors v and w is negative. We thus have proved

Lemma 2. *If of two independent vectors, v and w, one is non-positive and the other negative, then every vector of (v,w), i.e. the open first v,w-quadrant, is negative.*

Lemma 3. *If v and w are independent null vectors, then either every vector of (v, w) is null or every vector of (v,w) is negative.*

By Lemma 1, every vector of (v,w) is non-positive. Either every vector of (v, w) is null or there exists a negative vector, x, of (v,w). In the latter case, by Lemma 2, every vector of (v,x) and of (x,w) is negative. Since by Remark 1 of Section 3, every vector of (v,w) belongs either to (v,x) or to (x,w) or to the x-ray, every vector of (v,w) is negative.

Lemma 4. *If v and w are independent and v is degenerate, then every vector of the open half-plane $(v,w] + [w, -v)$ of the v, w-plane has the same sign as w.*

If w or any other vector of $(v,w] + [w, -v)$ is negative, then by Lemma 2 every vector in both quadrants is negative. If w is null, then by Lemma 1 every vector of $[v,w]$ and every vector of $[-v,w]$ is non-positive. By Lemma 3 none of these vectors is negative. Similarly, if any vector w' of the open half-plane is null, all vectors are null. Consequently, if w is positive, then every vector of the open half-plane is positive.

An obvious consequence of Lemma 4 is the following

Corollary. *If v and w are independent degenerate vectors, then every vector of the v, w-plane is degenerate.*

6. The Degenerate Part of Generalized Vector Spaces

The set, V_d, of all degenerate vectors of the vector space V is a subspace of V which we shall call *the degenerate part* of V. For if v is a degenerate vector, then obviously αv is degenerate for every α; and if v and w are degenerate, then by the Corollary of Lemma 4, the vector $v + w$ is degenerate.

Theorem 1. *Every generalized vector space, V, is the sum of a uniquely determined totally degenerate subspace, V_d, and a non-degenerate subspace V'. The latter can be chosen in such a way that V' and V_d are independent unless V is totally degenerate or non-degenerate. In the former case we have $V' = O^*$, in the latter case, $V_d = O^*$.*

Let V_d be the degenerate part of the vector space V. By the Lemma of Section 3, there exists a subspace V' such that $V = V_d + V'$ and $V_d \cdot V' = O^*$. The subspace V' is non-degenerate since V and V_d have only the vector o in common, and V_d contains all degenerate vectors.

7. The Non-Positive Part of a Vector Space

We shall call a subset, C, of a vector space a *cone* if

(*a*) C contains o and at least one vector besides o;

(*b*) for every vector, v, of C, except o, the v-ray is a subset of C.

We shall call the cone *convex* if

(*c*) for every two independent vectors, v and w, of C every vector of the first quadrant (v,w) belongs to C.

We shall call the cone *proper* if

(*d*) C does not contain two opposite vectors.

We shall refer to proper convex cones briefly as *cones*. We shall call the cone C ***open*** if every vector of C, except o, is an interior element of C. Here we define interior elements without reference to spherical neighbourhoods in the following way. The vector w is an *interior* element of the subset S of the vector space V if for every vector, v, of V there exists a positive number a (depending upon v) such that for every ϵ between 0 and a the vector $w + av$ belongs to S. We call a cone, C, *closed* if the set of all vectors not belonging to C is open. By the *boundary* of an open cone, C, we mean the set of all vectors w which do not belong to C while for some vector, v, of the vector space and every sufficiently small positive number ϵ the vector $w + \epsilon v$ does belong to C.

In terms of these concepts we can formulate the following

Theorem 2. *In a non-degenerate vectorspace, V, which is not definite, the set of all non-positive vectors is a closed cone. If V is indefinite, then the set consisting of o and all negative vectors is an open cone with the set of all null vectors as its boundary.*

Let V be a non-degenerate vector space which is not definite, that is to say, contains a non-positive vector $v \neq o$. Then the set, C, of all non-positive vectors is a cone since: (*a*) o is non-positive and, by assumption, V contains at least one vector $\neq o$; (*b*) every positive multiple of a non-positive vector is non-positive; (*c*) the convexity condition is satisfied by virtue of Lemma 1 of Section 5; condition (*d*) is satisfied because, by assumption, V is non-degenerate. The cone C is closed since, by virtue of the continuity of the norm, the set of all positive vectors is open.

Now let V be indefinite, that is, contain a negative vector. Then the set consisting of o and all negative vectors satisfies: conditions (*a*) and (*b*); moreover, by virtue of Lemma 2 of Section 5, condition (*c*); condition (*d*) since the opposite of a negative vector is positive. C is open by virtue of the continuity of the norm. Each null vector, v, belongs to the boundary of the open cone. For if x is a negative vector, then, for every positive a which is < 1, the vector $v + a(x - v)$ is negative. Hence v belongs to the boundary of the cone.

8. The Decomposition of Finite-Dimensional Spaces

Lemma. *If V^* is a finite-dimensional definite subspace and P a plane which is independent of V^* and such that $W = V^* + P$ is non-degenerate, then W contains a vector, w', such that $V^* + [w']$ is definite.* If $V^* = O^*$, the Lemma contends that *every non-degenerate plane contains a definite line.*

From the definiteness of V^* we deduce the following

Remark A. *If z is a non-positive vector of W, then for every vector v^* of V^* and every number $a > 0$, the vector $v^* - az$ is positive.*

For if $y = v^* - az$ were non-positive, then, since W is non-degenerate, y and z would be independent, and by Lemma 1, every vector of the first quadrant (y,z) would be non-positive whereas $v^* = y + az$ is positive.

From the finite dimensionality of V^* (and thus of W) we deduce a Remark B concerning a property of a particular subset, S, of W. *Only this single consequence* of the assumption that V^* has a finite dimension, say $k - 2$, will be used in the proof. If $x_1, \ldots, x_k$ are independent vectors of W, let S denote the set of all vectors $a_1x_1 + \ldots + a_kx_k$ for which $\Sigma a_i^2 = 1$.

Remark B. *Every sequence $s_1, s_2, \ldots$ of vectors of S contains a subsequence $s_{i_1}, s_{i_2}, \ldots$ for which a vector, s, of S exists such that*

$$\lim_{n\to\infty} | s_{i_n} | = | s |.$$

If, in particular, the kth components of the vectors $s_1, s_2, \ldots$ converge to 0, then s can be so chosen that its last component is 0. If the $(k-1)$th components of all vectors $s_1, s_2, \ldots$ are positive, then s can be so chosen that the $(k-1)$th component of s is non-negative.

By virtue of the compactness of the sphere $\Sigma a_i^2 = 1$ in the k-dimensional Cartesian space of the $(a_1, \ldots, a_k)$, the s_{i_n} can be so selected that if

$$s - s_{i_n} = a_{n,1}x_1 + \ldots + a_{n,k}x_k,$$

then as $n \to \infty$

$$\lim a_{n,1} = \lim a_{n,2} = \ldots = \lim a_{n,k} = 0.$$

Hence Remark B is a consequence of the finite-dimensional continuity of the norm.

If every vector of W is positive, then the proposition of the Lemma holds. We thus can assume that W contains a non-positive vector, v. Its opposite, the vector $-v$, is positive. By Remark A, for every v^* of V^* and every $a > 0$, the vector $v^* - av$ is positive. W contains a vector, w, such that $[w]$ and $V^* + [v]$ are independent. Now we prove:

There exists a $\beta \geq 0$ with the following property P. If $a > \beta$, then for every v^* of V^*, the vector $v^* - av + w$ is positive.

We assume that no number $\beta \geq 0$ has the property P and derive a contradiction. By the assumption, for every n there exists a number $a_n > n$ and a vector, v^*_n of V^* such that

$$y_n = v^*_n - a_nv + w$$

is non-positive. Let $x_1, \ldots, x_{k-2}$ be independent vectors of V^* such that

$$v^*_n = a_{n,1}x_1 + \ldots + a_{n,k-2}x_{k-2}.$$

We set

$$x_{k-1} = -v \text{ and } a_{n,k-1} = a_n,$$
$$x_k = w \text{ and } a_{n,k} = 1,$$
$$\left[\sum_i (a_{n,i})^2\right]^{\frac{1}{2}} = \gamma_n > 0$$

Since $a_n > n$ we have $\lim \gamma_n = \infty$. If we set

$$s_n = \frac{1}{\gamma_n} y_n ,$$

then by Remark B there exists a subsequence $s_{i_1}, s_{i_2}, \ldots$ and a vector s of S such that

$$\lim |s_{i_n}| = |s|.$$

Since $\lim \gamma_n = \infty$ and $a_{n,k} = 1$ and $a_{n,k-1} > 0$ for every n, we see that s can be chosen as a vector of the form $v^* - av$ where v^* is in V^* and $a \geq 0$. Thus s is positive. This is a contradiction since the s_n, and, in particular, the s_{i_n} are non-positive. (The y_n are non-positive, the $\gamma_n > 0$).

Having established the existence of a $\beta \geq 0$ with the property P, we call β_0 the g.l.b. of all $\beta \geq 0$ with the property P. Two cases are possible.

First Case. $\beta_0 > 0$. Then from the definition of β_0 it follows[1] that there exists a non-positive vector (and thus clearly a null vector) $w_0 = v^*_0 - \beta_0 v + w$ where v^*_0 belongs to V^*. Now if for a vector v^* of V^* and $\kappa > 0$ the vector $w'_\kappa = v^* - (\kappa - \beta_0)v - w$ were non-positive, then since $V^* + P$ is non-degenerate, the vectors w_0 and w' would be independent, and all vectors of the first quadrant (w_0, w') would be non-positive. But this is not the case since $w_0 + w'$, that is, $(v^*_0 + v^*) - \kappa v$ is positive. Hence all vectors w'_κ are positive. Since w_0 is non-positive, by Remark A, for every vector v^* of V^* and every $a > 0$, the vector

$$a\left(\frac{1}{a} v^* - v^*_0\right) + a\beta_0 v - aw$$

is positive. Hence for every v^*_1 of V^*, the vector $v^*_1 + \beta_0 v - w$ is positive. In particular, $-v^*_0 + \beta_0 v - w$ is positive. Since its opposite, the vector w_0, is non-positive, by Remark A all vectors $v^* - aw_0$ or, which is equivalent, all vectors

$$v^* + a(\beta_0 v - w) \text{ for } v^* \text{ in } V^* \text{ and } a > 0$$

are positive. From this fact it follows [exactly as from the positivity of all vectors $v^* - av$ we derived the existence of a $\beta \geq 0$ with the property P] that there exists a $\gamma \geq 0$ such that for each v^* of V^* and every $a > \gamma$ the vector

$$v_a = v^* + a(\beta_0 v - w) + w$$

is positive. No matter how we determine $a > \gamma$, for every $\pi > 0$ the vectors

$$\pi[v^* + a(\beta_0 v - w) + w]$$

[1] If we use an argument similar to the one which yielded the existence of s.

or, which is equivalent, the vectors

$$v^* + \pi[\alpha(\beta_0 v - w) + w]$$

are positive for every v^* of V^*. Now if $\alpha > 1$, then for every v^* of V^* the vector

$$v^* - \alpha(\beta_0 v - w) - w = (\alpha - 1)\left[\frac{1}{\alpha - 1}v^* + w - \frac{\alpha}{\alpha - 1}\beta_0 v\right]$$

is positive by definition of β_0. Hence every vector

$$v^* - \pi[\alpha(\beta_0 v - w) + w]$$

is positive. Thus if we set, for instance,

$$w' = (2 + \gamma)(\beta_0 v - w) + w$$

then $v^* + \alpha w'$ is positive for every v^* in V^* and every α, that is to say, the subspace $V^* + [w']$ is definite.

Second Case. $\beta_0 = 0$. Then we first let w and v play the roles of $-v$ and w, respectively, and arrive at a γ_0 such that for $\alpha > \gamma_0$

$$v^* + \alpha w$$

is positive while one vector

$$w_1 = v^* + \gamma_0 w$$

is non-positive (thus obviously null). We let now this vector play the role of w_0 and exactly as before arrive at a vector w'' such that $V^* + [w'']$ is definite. This completes the proof of the Lemma.

Now let V be a finite-dimensional non-degenerate vector space. If V is a plane, then, by the Lemma, V contains a definite line. By induction we build an increasing chain of definite subspaces which, by the Lemma, can be continued as long as there exists a plane which is independent of the subspace. This is the case until we have reached a definite subspace, V', whose dimension is that of V minus 1. If V itself is not definite we can represent V as the sum of V' and the line of any vector not contained in V'. As such a vector we may use any non-positive vector v'. The opposite of v' is positive. We thus have proved the following

Theorem 3. *Every finite-dimensional non-degenerate space which is not definite is the sum of a definite subspace and the v'-line of any non-positive vector v' which is $\neq o$.*

Since every indefinite vector space obviously contains a nullvector besides o, and in a non-degenerate space the line of such a vector is semi-definite, we obtain the following

Corollary. *Each non-degenerate finite-dimensional vector space which is non definite is the sum of a definite subspace and a single semi-definite line.*

Combining Theorems 1 and 3 we can say

Each finite-dimensional vector space, V, can be represented as the sum of the following parts:

(1) *a uniquely determined totally degenerate subspace (which is O^* if and only if V is non-degenerate);*

(2) *a definite subspace (which may be O^*); to which, if and only if V con-*

tains a non-positive vector whose opposite is positive, we have to add a third subspace, namely,

(3) *a single line for which we may choose the v-line of any non-positive vector.*

9. A Survey of All Finite-Dimensional Vector Spaces

The space O^*, and only this space, may be considered as 0-dimensional. There are four types of 1-dimensional spaces ("lines"): definite, semi-definite and indefinite lines (which are non-degenerate) and totally degenerate lines. By induction we see that there are the following types of n-dimensional spaces:

I. *Non-Degenerate Spaces.*

1. *Definite Spaces.*
2. *Semi-Definite Spaces.* We shall classify them by calling the dimension of the closed cone of null vectors of a space, its *degree.*
3. *Indefinite Spaces* with an n-dimensional cone of negative vectors whose boundary consists of the null vectors.

II. *Degenerate Spaces* which are the sum of a non-degenerate space of a dimension $< n$ (which we shall call the *rank* of the space) and a totally degenerate space.

In subsequent papers we shall refer to the definite, semi-definite and indefinite spaces as *elliptic, parabolic,* and *hyperbolic* spaces, respectively. The parabolic spaces of minimum degree, 1, will be called *properly* parabolic. We shall see that in a finite-dimensional vector space every closed cone, for a properly chosen definition of the norm, is the cone of the null vectors of a parabolic space; and that every open cone may be the cone of the negative vectors of an indefinite space. Hence every hyperbolic space can be made parabolic of maximum degree by redefining the norm of each negative vector as 0, while every parabolic space of maximum degree can be made hyperbolic by associating with the interior null vectors proper negative norms.

References

[1]S. Banach, *Fund. Math.*, vol. 3 (1922), 133-181; H. Hahn, *Monatshefte Math. Phys.*, vol. 32 (1922), 1-81; N. Wiener, *Bull. Soc. Math. France*, vol. 150 (1922), 124-134.

[2]Cf., in particular, "Die metrische Methode in der Variationsrechnung," *Ergebn. mathem. Kolloquiums*, vol. 8 (1937), 1-32; two notes in the *Proc. Nat. Acad. Sci.*, vol. 23 (1937), 246 and vol. 25 (1939), 474; and the lecture "Analysis and Metric Geometry" in The Rice Institute Pamphlet, vol. 27 (1940), 1-40.

[3]*Ergebn. mathem. Kolloquiums*, vol. 8 (1937), 32.

[4]Cf., in particular, Pauc's résumé in the pamphlet "Les méthodes directes en Calcul des Variations," (Paris, Herman, 1941).

[5]*Rend. della Acc. Naz. Linc.*, vol. 26 (1937).

[6]*Ergebn. math. Kolloquiums*, vol. 8 (1937), 25 and Alt, loc. cit.[3]

Illinois Institute of Technology
Chicago

GROUPS WITH REPRESENTATIONS OF BOUNDED DEGREE

IRVING KAPLANSKY

1. Introduction. Let G be a compact group. According to the celebrated theorem of Peter-Weyl there exists a complete set of finite-dimensional irreducible unitary representations of G, the completeness meaning that for any group element other than the identity there is a representation sending it into a matrix other than the unit matrix. If G is commutative, the representations are necessarily one-dimensional. It is an immediate consequence of the Peter-Weyl theorem that the converse also holds: if every representation is one-dimensional, G is commutative. The main theorem in the present paper is a generalization of this result to the case where the representations have bounded degree. We may illustrate by stating the next simplest case. The representations are one- or two-dimensional if and only if G satisfies the following condition: for any 4 elements of G the 12 $(= 4!/2)$ products obtained from even permutations can be paired off in equal pairs with the 12 products obtained from odd permutations. The general result is stated in Theorem 3.

Such groups exist: for example, the group extension of an abelian group by a finite group (Theorem 1). On the other hand, if such a group is connected it is abelian (Theorem 2).

In §§ **2**, **3** we present some preliminary remarks on matrices and groups, and in § **4** we review some facts on group representations needed for the extension from the compact to the locally compact case. In § **5** the main theorems appear, and in § **6** a connection with a theorem due to Halmos is described.

2. Matrix identities.[1] For elements $x_1, \ldots, x_r$ in a ring we shall write

$$[x_1, \ldots, x_r] = \sum \pm x_{\pi(1)} \ldots x_{\pi(r)}$$

where the sum runs over all permutations π and the plus or minus sign is prefixed according as π is even or odd.

LEMMA 1. *In any algebra A of order $k-1$ we have $[x_1, \ldots, x_k] = 0$ for all $x_i \epsilon A$.*

Proof. Since the relation in question is multilinear, it need only be proved when $x_1, \ldots, x_k$ are basis elements. In that case at least one repetition occurs, and consequently a transposition can be performed which leaves $[x_1, \ldots, x_k]$ unchanged. Hence

$$[x_1, \ldots, x_k] = -[x_1, \ldots, x_k] = 0.$$

Received February 23, 1948.

[1] I am greatly indebted to E. R. Kolchin for the contents of 2.

(Formally this argument is invalid for characteristic 2, but the result is still correct and may be proved by the usual device of a reduction mod 2.)

We may apply Lemma 1 to the special case where A is the algebra of n by n matrices over a field. We shall write $r(n)$ for the smallest integer such that $[x_1, \ldots, x_r] = 0$ for all n by n matrices.[2] By Lemma 1 we have $r(n) \leq n^2 + 1$.

The following argument gives a lower bound for r. Write $t = r(n-1) - 1$. Suppose $x_1, \ldots, x_t$ are $n-1$ by $n-1$ matrices with $[x_1, \ldots, x_t] \neq 0$, and we may suppose to be explicit that $[x_1, \ldots, x_t]$ contains a non-zero term in e_{kk}, where $\{e_{ij}\}$ denote the usual matrix units. Embed the matrix x_i in an n by n matrix y_i by adjoining a row and column of zeros. Then it is evident that

$$[y_1, \ldots, y_t, e_{kn}, e_{nn}] \neq 0.$$

This proves the following result.

Lemma 2. $r(n) \geq r(n-1) + 2$.

It is clear that $r(1) = 2$ and by Lemma 2 we deduce the lower bound $r(n) \geq 2n$. For $n = 2$, $r(2)$ is in fact precisely 4. This apparently exhausts the known facts concerning $r(n)$.

3. A certain class of groups. Let us say that a group G satisfies the condition $P_n (n \geq 2)$ if the following is true: for any n elements in G the set of $n!/2$ products obtained from even permutations coincides with the $n!/2$ products obtained from odd permutations. It should be noted that it is not asserted that there is a fixed way of carrying out the pairing once for all; the particular correspondence presumably depends upon the particular n elements in question.

It is fairly evident that P_k implies P_{k+1}. P_2 simply asserts commutativity, and so does P_3 as can be seen by taking one of the three elements to be the identity. Starting at $k = 4$ there exist non-abelian groups satisfying P_k; for example, the symmetric group on three elements satisfies P_4. The following theorem provides us with a substantial class of such groups.

Theorem 1. *A group extension of an abelian group by a finite group of order n satisfies P_{n^2+1}.*

Proof. We suppose that G is abelian, H of order n, and $K/G \cong H$. Choose fixed representatives $k_1, \ldots, k_n \epsilon K$ for the cosets of K mod G. Every element of K can be uniquely written gk_i, $g \epsilon G$. Let b be a product of $n^2 + 1$ elements of K. In such a product some k, say k_1, must be repeated at least $n + 1$ times. Let gk_1 be the element appearing at the first occurrence of k_1, and $g'k_1$ one of the later occurrences. Write x for the product of the k's intervening between these two instances of k_1. The interchange of the pair gk_1 and $g'k_1$ will leave b unchanged provided that k_1x lies in G. Since we have $n + 1$ or more occurrences of k_1 and only n cosets of K mod G, it will have to happen at least once that an interchange of two of the terms comprising b leaves b unchanged.

---

[2] It is conceivable that $r(n)$ depends on the coefficient field, or rather on the characteristic of the latter. To be explicit, one may take the characteristic 0 case throughout the paper.

We now specifically pick out the *first* element in the product b whose interchange with a later element is legal. In all the $(n^2+1)!$ permutations we do the same thing, and thus set up a one-to-one correspondence between the even permutations and the odd permutations. This proves Theorem 1.

Theorem 1 does not give the best possible result. Indeed we shall show below that a group extension of an abelian group by a group of order n actually satisfies $P_{s(n)}$, where

$$(1) \qquad s(n) = r(n) \text{ for } r(n) \text{ even}$$
$$= r(n)+1 \text{ for } r(n) \text{ odd},$$

and $r(n)$ is the integer defined in § 2. Thus for $n = 2$ we get P_4 instead of the P_5 of Theorem 1. However I am unable to prove this refinement without the detour to group representations and Banach algebras.

We shall conclude this section by showing that there are no connected non-abelian groups of the kind under discussion. Actually we prove a (formally at least) stronger result, in order to carry through an induction.

THEOREM 2. *Let G be a connected topological group having for some fixed $n \geqq 2$ the following property: any product $a_1a_2 \ldots a_n$ is equal to a proper permutation. Then G is abelian.*

Proof. We shall show that G has the same property for $n-1$ and hence finally reach $n = 2$. Let then $a_1, \ldots, a_{n-1}$ be elements in G. For any b in G the product $a_1 \ldots a_{n-1}b$ must be equal to a proper permutation. We may suppose that there is a neighbourhood U of the identity such that for b in U the proper permutation in question keeps the order of $a_1, \ldots, a_{n-1}$ fixed; for otherwise we can take the limit as b approaches the identity and conclude that $a_1 \ldots a_{n-1}$ equals a proper permutation. Thus for each b in U we have one of the $n-1$ possible equations

$$a_1 \ldots a_{n-1}b = a_1 \ldots a_i b a_{i+1} \ldots a_{n-1} (i = 0, \ldots, n-2).$$

The ith equation asserts that b commutes with $a_{i+1} \ldots a_{n-1}$ and hence is valid in a closed set. Thus U is covered by a finite number of these closed sets, and one of them must have a non-void interior. This says that the centralizer of $a_{i+1} \ldots a_{n-1}$ is open. Since G is connected, this centralizer must be all of G and hence $c_{i+1} \ldots a_{n-1}$ is in the centre. For $i \geqq 1$ this yields the desired result obviously, while for $i = 0$ the assertion that $a_1 \ldots a_{n-1}$ is in the centre implies

$$a_1a_2 \ldots a_{n-1} = a_2 \ldots a_{n-1}a_1.$$

COROLLARY. *If a connected topological group satisfies P_n it is abelian.*

4. Group representations. In order to formulate our main theorem for a locally compact group G, it would not suffice to assume that the finite-dimensional irreducible unitary representations of G have bounded degree; for there exist groups (e.g. the Lorentz group) for which the only finite-dimensional unitary representation is the trivial one. Thus we must impose a further condition which will entail the existence of a respectable number of finite-dimensional representations. For our purposes a convenient hypothesis of

this kind can be formulated in terms of the representations introduced by Segal [4]. We devote this section to a brief statement of the necessary facts.

Let A denote the L_1-algebra of the locally compact group G, that is, the algebra of all complex-valued functions summable with respect to the left Haar measure of G, with convolution as multiplication:

$$fg(x) = \int f(y)g(y^{-1}x)dy.$$

Let E be the algebra of bounded operators on a Banach space. A B-representation [4, p. 79] of G is a multiplicative homomorphism T of G into E which sends the identity into the unit operator, is continuous in the strong topology of E, and is such that $||T(a)||$ is bounded for $a \epsilon G$. A B-representation is irreducible if it admits no proper closed invariant subspaces. Irreducible B-representations may be constructed as follows. Let M be a regular maximal left ideal in A, and associate with $a \epsilon G$ the operator $T_a: u + K \rightarrow u_a + K$, where $u_a(x) = u(a^{-1}x)$. We shall call these representations *primitive*, a designation suggested by the fact that the extension of the representation to A has as its kernel the ideal P consisting of all x with $xA \leqq M$; P is a primitive ideal in the sense of Jacobson [2]. Conversely every primitive ideal in A is associated in this fashion with at least one primitive representation of G.

The following facts are known: (1) all primitive representations are irreducible, (2) any irreducible finite-dimensional unitary representation is similar to a primitive representation, (3) if G is compact or abelian, all primitive representations of G are finite-dimensional. It is an open question whether every irreducible B-representation is similar to a primitive representation.

5. Main theorem. In terms of the concepts introduced in the previous sections, the principal result can be stated as follows.

THEOREM 3. *The following two statements are equivalent for a unimodular locally compact group G:*[3]

(*a*) *All primitive representations of G are finite-dimensional and of degree at most n,*

(*b*) *G satisfies the condition $P_{s(n)}$, where $s(n)$ is defined by* (1).

It is to be observed that if G is compact, the theorem simplifies perceptibly: compact groups are unimodular, and their primitive representations are automatically finite-dimensional.

Proof. Suppose that (*a*) holds. Then it follows virtually from the definition of the primitive representations that for every primitive ideal P in $A = L_1(G)$, $A - P$ is finite-dimensional and is in fact a total matrix algebra of degree at most n. Hence $A - P$ satisfies the identity $[x_1, \ldots, x_k] = 0$ for $k = r(n)$ and *a fortiori* for $k = s = s(n)$. Now the intersection of the primitive ideals of A is 0: this is a consequence of the semi-simplicity of A: [4, Th. 1.5]

[3]A group is unimodular if its right and left Haar measures coincide—cf. [5, p. 39].

and [2, Th. 25]. Hence $[f_1, \ldots, f_s] = 0$ holds for all $f_i \epsilon A$. The s-fold convolution of functions is given by

$$(2) \qquad f_1 \ldots f_s(x)$$
$$= \int \ldots \int f_1(y_1)f_2(y_2) \ldots f_{s-1}(y_{s-1})f_s[(y_1 \ldots y_{s-1})^{-1}x]dy_1 \ldots dy_{s-1}.$$

We shall now study the effect of a permutation π on $f_1 \ldots f_s$. If π does not involve the letter s, its effect on (2) may be described as carrying out π on the y's in $(y_1 \ldots y_{s-1})^{-1}$, and otherwise leaving the right side of (2) unchanged. Next we try the case $\pi = (is)$. We carry out the interchange of f_i and f_s in (2) and then replace y_i by

$$(3) \qquad (y_1 \ldots y_{i-1})^{-1}xy_i^{-1}(y_{i+1} \ldots y_{s-1})^{-1},$$

(a legal change of variable in view of the assumed unimodularity of G). This replaces

$$(4) \qquad (y_1 \ldots y_{s-1})^{-1}x$$

by y_i and so finally gives us the right side of (2) changed by the substitution of (3) for (4). In view of the fact that s is even, it can be verified that the permutation (4) $\rightarrow$ (3) is odd.

The general permutation π which does involve s can be written uniquely as $\pi = (is)\pi_1$, where π_1 is independent of s. The effect of π on the right side of (2) can thus be described as changing the argument of f_s by the permutation (4) $\rightarrow$ (3), followed by the permutation π_1 on the y's. This is a one-to-one correspondence: given the induced permutation on the argument of f_s we can unambiguously reconstruct π; for the position of x (in the ith place) gives us the portion (is), and the position of the y's then determines π_1. Moreover, the correspondence preserves the parity of π, as we have seen.

We may summarize as follows. We have

$$(5) \qquad \int \ldots \int f_1(y_1)f_2(y_2) \ldots f_{s-1}(y_{s-1})Zdy_1 \ldots dy_{s-1} = 0,$$

where we have written Z for

$$Z = \sum_j \pm f_s(z_j),$$

z_j being the general permutation of (4), and the plus or minus sign being taken according to the parity of the permutation z_j. Since (5) holds for all f_i in A and in particular for all continuous f_i in A we deduce that $Z = 0$ for all continuous f_s in A. Since a continuous function in A can take arbitrary values at any finite subset of G, we conclude that G must satisfy the condition P_s.

We now proceed to the proof of the converse. Suppose that (b) holds. Then the computation above is reversible to the point where we have $[f_1, \ldots, f_s] = 0$ for $f_i \epsilon A$. Let P be a primitive ideal in A; the identity $[x_1, \ldots, x_s] = 0$ is of course inherited by $A - P$. Theorem 1 of [3] asserts that a primitive algebra satisfying a polynomial identity is finite-dimensional over its center. Hence $A - P$ is finite-dimensional over its centre C. By the Gelfand-Mazur theorem on normed fields, C is just the complex numbers. Hence $A - P$ is an algebra

of finite order over the complex numbers, and is indeed a full matrix algebra. As for the degree of these matrices, it cannot exceed n; for by Lemma 2, $s \leqq r(n)+1 < r(n+1)$ and consequently matrices of degree $n+1$ fail to satisfy $[x_1, \ldots, x_s] = 0$. This shows that the primitive representations of G are finite-dimensional and of degree at most n, and concludes the proof of Theorem 3.

The criterion provided by Theorem 3 is in many cases easy to apply. For example, let G and H be unimodular locally compact groups whose primitive representations have bounded degree; then from Theorem 3 it follows that the same is true for $G \times H$, any unimodular homomorphic image of G, and any closed unimodular subgroup of G. Also the following result is a corollary of Theorems 2 and 3.

COROLLARY. *Let G be a connected unimodular locally compact group whose primitive representations are finite-dimensional and of bounded degree. Then G is abelian.*

This corollary may be derived in another way which we shall now describe. Let G be a connected locally compact group which is maximally almost-periodic, that is, G has a complete set of finite-dimensional unitary representations. (This hypothesis is weaker than the assumption that the primitive representations of G are finite-dimensional.) By a theorem of Freudenthal [5, p. 129], G is the direct product of a compact group and a finite number of copies of the additive group of real numbers. The question as to when the irreducible unitary representations are of bounded degree is thereby reduced to the compact case; and by considering the images under the representations, we further reduce to the case of a compact Lie group. In fact, our problem becomes precisely the following: prove that a connected compact simple Lie group possesses irreducible representations of arbitrarily high degree. That this is in fact the case follows from known classical results.

The corresponding theorem for Lie algebras asserts that a simple Lie algebra has irreducible representations of arbitrarily high degree. In this form, the theorem has recently been given a purely algebraic proof by Harish-Chandra [6]. It is perhaps worth remarking that, by standard devices, the theorem on Lie algebras can conversely be derived from the group theorem.

We return to the study of the group K of Theorem 1, and shall derive the purely group-theoretic fact that K satisfies $P_{s(n)}$. We give K the discrete topology, which assures its local compactness and unimodularity. Then by Theorems 1 and 3 we have that the primitive representations are finite-dimensional and of bounded degree. Theorems 1 and 3 also yield a bound for the degree in question, but a better bound can be obtained by a simple direct argument. In fact we assert that any finite-dimensional irreducible unitary representation T of K is of degree at most n. For the induced representation of G decomposes into one-dimensional representations, since G is abelian. Let α be a non-zero vector in one of these G-invariant one-dimensional subspaces. Then for $g \epsilon G$, $\alpha T(g)$ is a multiple of α. Using the notation of the

proof of Theorem 1, we deduce that the invariant subspace generated by a is spanned by $aT(k_1), \ldots, aT(k_n)$ and hence is at most n-dimensional, as desired. Quotation of Theorem 3 proves that K satisfies $P_{s(n)}$.

We shall conclude this section with a variant of Theorem 3:

THEOREM 4. *The following two statements are equivalent for a unimodular locally compact group G*:

(*a*) *The primitive representations of G are finite-dimensional and of bounded degree.*

(*b*) *The L_1-algebra of G satisfies a polynomial identity.*

Proof. The proof coincides with the corresponding portions of the proof of Theorem 3, except for the following remark. In proving that (*b*) implies (*a*) we take a primitive ideal P in $A = L_1(G)$ and quote Theorem 1 of [3] to sustain the claim that $A - P$ is finite-dimensional. But more than that: Theorem 1 of [3] shows that the dimension of $A-P$ has a fixed upper bound depending only on the polynomial identity in question (cf. remark (*b*) on p. 580 of [3]). The rest of the proof proceeds unchanged.

6. A theorem of Halmos. The study of groups with bounded representations arose in connection with an attempt to generalize a theorem of Halmos, which we shall now describe. Let G be a compact group and S a continuous automorphism of G. The uniqueness of Haar measure shows that S induces a measure preserving transformation on G, which in turn induces a unitary operator $U: f \rightarrow f^S$ on $L_2(G)$. We say that S is ergodic if the only solutions of $f^S = f$ are constant. In [1] Halmos studied the case where G is commutative, and showed (Th. 3) that if S is ergodic, the spectral type of U is entirely determined by the cardinal number of the character group of G. We refer the reader to [1] for the precise result.

Now let G be a compact group which is not necessarily commutative. The automorphism S induces in a natural way a permutation π_S of the irreducible representations of G. This permutation leaves the trivial representation ψ fixed (ψ sends every element into the matrix (1)). The analogue of Halmos's [1, Th. 1] is now valid: S is ergodic if and only if π_S has no finite orbits other than ψ. The proof is virtually the same as that given by Halmos: one uses the coordinates of irreducible representations in place of characters.

Supposing that S is ergodic, we can now proceed to discuss the spectral type of U. Of course $U(\psi) = \psi$. By appropriate choice of the remaining coordinates of irreducible representations, which together with ψ form an orthonormal base of $L_2(G)$, we can arrange them in a double array $\phi_{i,j}$ such that $U(\phi_{i,j}) = U(\phi_{i,j+1})$. Here the index j runs over all integers, and the index i over the orbits of π_S. If we let c denote the number of orbits in question, we have proved Halmos's [1, Th. 3] except for the assertion that c is infinite.[4] If the

[4] If there are an uncountable number of irreducible representations, it is clear that c is infinite (and equal to that number). Thus further discussion is really needed only for the case of a countable number of irreducible representations.

representations have unbounded degree, then it is clear that c is infinite, for the permutation π_S necessarily preserves degree. At the other extreme, if all the representations are one-dimensional (G commutative), Halmos provided a group-theoretic argument on the character group to show that c is infinite [1, Th. 2]. There remains the case of representations of bounded degree, where it would be necessary to generalize suitably Halmos's argument. I have been unable to supply such an argument, but possibly the results in this paper will point the way toward the completion of this problem.

Postscript (*December* 1, 1948)

Since this manuscript was completed, a paper by F. W. Levi has appeared: "On Skew Fields of a Given Degree," *J. Indian Math. Soc.*, vol. II (1947), 85-88. Reference is made there to a paper to be published in the *Mathematische Annalen*. In the notation of §2, this latter paper proves (among other things) that $r(n)$ is even and $r(3) = 6$. The distinction between $r(n)$ and $s(n)$ may therefore be suppressed.

References

[1] P. R. Halmos, "On Automorphisms of Compact Groups," *Bull. Amer. Math. Soc.*, vol. 49 (1943), 619-624.

[2] N. Jacobson, "The Radical and Semi-simplicity for Arbitrary Rings," *Amer. J. of Math.*, vol. 67 (1945), 300-320.

[3] I. Kaplansky, "Rings with a Polynomial Identity," *Bull. Amer. Math. Soc.*, vol. 54 (1948), 575-580.

[4] I. E. Segal, "The Group Algebra of a Locally Compact Group," *Trans. Amer. Math. Soc.*, vol. 61 (1947), 69-105.

[5] A. Weil, "L'Intégration dans les groupes topologiques et ses applications," Actualités Scientifiques et Industrielles, no. 869, Paris, 1938.

[6] Harish-Chandra, "On Representations of Lie Algebras," to appear in the Annals of Mathematics.

University of Chicago